

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:32:52 ; Search time 1822.81 Seconds
(without alignments)
12358.072 Million cell updates/sec

Title: US-09-591-466c-1

Perfect score: 1669

Sequence: 1 gaattcgccgcgcctgaga.....actttggcgcgcgaattc 1669

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*

```

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_esti.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pln.*
16: em_gss_vrt.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	719.8	43.1	768	10	BG594788
2	694	41.6	724	10	BI179808
3	638.6	38.3	661	10	BG889872
4	631.4	37.8	642	10	BG592816
5	615.8	36.9	643	9	A1486771
6	598.6	35.9	673	9	AW979500
7	586.2	35.1	659	9	AW735827
8	549	32.9	574	9	AW931180
9	479.6	28.7	602	10	BG131020
10	441.6	26.5	575	10	BG887316
11	431.6	25.9	681	9	A1725875
12	423.8	25.4	528	9	A1897249
13	408	24.4	488	10	BG88733
14	395	23.7	530	9	AW618699
15	392.4	23.5	629	10	BE426174
16	384.6	23.0	507	9	A1773401
17	380.4	22.8	664	10	BE805530

18	363.4	21.8	591	10	BG409413
19	327.2	19.6	603	10	BE516817
20	312.6	18.7	506	10	BG726055
21	288.6	17.3	488	10	BG508571
22	284	17.0	525	10	BM177681
23	277.4	16.6	418	10	BF066219
24	261.4	15.7	472	10	BG051233
25	259.6	15.6	701	10	BJ170544
26	256.2	15.4	635	10	BF145860
27	238.6	14.3	691	10	BE821775
28	222.6	13.3	588	10	BE404993
29	216.4	13.0	495	10	BG154089
30	212	12.7	388	10	BE325650
31	212	12.7	538	10	BE920532
32	207.8	12.5	227	10	BE461505
33	205.6	12.3	460	10	BG043733
34	200.8	12.0	2681	11	AK004760
35	189.6	11.4	448	10	BG159494
36	188.4	11.3	464	9	AW695874
37	174	10.4	386	10	BG046010
38	167.4	10.0	627	10	BM359759
39	167.4	10.0	894	10	BG680531
40	166.2	10.0	629	10	BG444140
41	163.2	9.8	801	10	BI152075
42	162.6	9.7	546	9	AW670629
43	161	9.6	889	10	BI415357
44	158.6	9.5	756	10	C96708
45	156.6	9.4	677	10	BE289472

ALIGNMENTS

RESULT	1
BG594788	
LOCUS	EST493466 cSTS Solanum tuberosum cDNA clone CSTS8020 5' sequence,
DEFINITION	768 bp mRNA linear EST 12-APR-2001
ACCESSION	BG594788
VERSION	EST.
KEYWORDS	potato.
SOURCE	Solanum tuberosum
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE	1 (bases 1 to 768)
AUTHORS	van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE	Generations of ESTs from sprouting potato eyes
JOURNAL	Unpublished (2000)
COMMENT	Contact: Cathy Ronning The Institute for Genomic Research For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com Seq primer: M13F-R.
FEATURES	Location/Qualifiers
source	1..768 /organism="Solanum tuberosum" /cultivar="Kennebec" /db_xref="taxon:4113" /clone="CSTS8020" /clone.lib="cSTS" /tissue_type="sprouting eyes from tubers" /dev_stage="12-14 weeks post harvest" /lab_host="SOLR" /note="Vector: pBluescript SK(-); Site:1: EcoRI; Site:2: XhoI; Various sizes of sprouting eyes (2mm to 15mm) were taken from tubers. The tubers were incubated at 26C in the dark for 2-3 weeks prior to sprouting. The eyes were frozen in liquid nitrogen immediately upon removal from tubers."
BASE COUNT	220 a 149 c 181 g 218 t

ORIGIN

Query Match 43.1%; Score 719.8; DB 10; Length 768;
Best Local Similarity 99.0%; Pred. No. 2.4e-186;
Matches 724; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 15 ctgagaacccctgaattcaatttcgcatcttgccagatgagaggaacaagtcttgct 74
Db CTGAGAAACCCTCGAATTCATTTTCGATTTGGCAGAGATGAGAGGAACAAGTTTGTCT 95

QY 75 ttgatttacggtacactctctgcgtgctgctgcgtcttcattacacagatgcgc 134
Db TTGATTTACGGTACCTTCTCGCTGCTGCTCTCGCTTCATCTACATACAGATGCGGC 155

QY 135 ttctcgacacagtcagaatatagaccgcttgcgtgcgaattgaagcaaaaatc 194
Db TTTTCGGCACACATCAGATATGATACACCCCTTGTCTGCAATTTAAGCAGAAAATC 215

QY 195 attgacaagtcagaccagattgcttattgacaagattaccagcagcaaggaagtag 254
Db ATTGTACAAGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGGAAGACTAG 275

QY 255 tagctcttgaagaacaataaagcatcagaccagagtgccggcaattaaaggctcttg 314
Db TAGCTCTTGAAGAACAATGAAGCGTCAGGACCAGGAGTGCCGGCAATTAAGGGCTCTTG 335

QY 315 ttccagatcttgaagtaaggccataaaaaagtttaacgagagtgatgcagatgccagtgg 374
Db TTCAGGATCTTGAAGTAAGGGCATAAAAAAGTTAATCGGAGATGTCAGATGCCAGTGG 395

QY 375 cagctgtagttgttatgcttgcaagtcgactgactaccctggagaggagactattaaatcca 434
Db CAGCTGTAGTGTATTGGCTTGCGAGTCGTCTGACTACCTCGAGAGGACTATTAAATCCA 455

QY 435 tcttaataaccacaacatctgttgcatacaaaatctcttcttccatccagatggat 494
Db TCTTAAATACCAACATCTTGCATCAAAATATCTCTTTTCATATCCAGATGGAT 515

QY 495 caaatctgatgtaagaaagcttgcgttgatgctatgctgctgagcgtatgcagcact 554
Db CAAATCCTGATGTAAAGAAAGCTTGCTTTGAGTATGATCAGCTGACGTATATCAGCACT 575

QY 555 tgattatgaacctgtgcatactgaaagaccagaggaactggtgcatactacaagattg 614
Db TGAATTTGAACCTGTGCATCTCAAGACCAGGGAACTGGTTGCATACTACAAGATTG 635

QY 615 cagctcattacaagtgggcattggatcagctgtttcacaagcataattttgacgctgita 674
Db CAGCTCATTTACAAGTGGGCATTGGATCAGCTGTTTCACAAGCATAATTTTAGCCGTGTTA 695

QY 675 tcatactagaagatgatagaaattgctgctgatttttttgactattttgagggctggag 734
Db TCATACTAGAAGATGATATGGAATTTGCTGCTGATTTTGTGACTATTTTAGGGCTGAG 755

QY 735 ctactctctt 745
Db CTACTCTCTT 766

RESULT 2

LOCUS B1179808
DEFINITION B1179808 724 bp mRNA linear EST 09-JUL-2001
EST520753 cSTE Solanum tuberosum cDNA clone cSTE20K20 5' sequence,
mRNA sequence.
ACCESSION B1179808
VERSION B1179808.1 GI:14645619
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE
AUTHORS

1 (bases 1 to 724)
van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,
Chiemingo,A., Bougfri,O., Buell,C.R., Ronning,C., Tanksley,S. and
Baker,B.

TITLE
JOURNAL
COMMENT

Generation of ESTs from in vitro grown microtubers
Unpublished (2001)
Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
source

Location/Qualifiers
1..724
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTE20K20"
/clone_lib="cSTE"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue supplied by Christian Bachem and Richard
Visser (Department of Plant Breeding, Wageningen
University, The Netherlands). The cSTA libraries will
attempt to capture the induction and initiation/initial
growth of the tuber in an in vitro system as described in
Bachem et al. (Plant Journal, 1996). Small microtubers
develop from axillary buds attached to stem explants when
placed on a high sucrose medium (10%). Visible
morphological changes occur synchronously at day five in
the axillary buds. The first library, cSTA (1-20) consists
of axillary buds harvested on days 1-3. This targets
those genes involved in induction of the microtubers. The
following libraries, cSTA (21-40) and cSTA (41-60),
capture genes involved in tuber initiation and outgrowth.
This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT
ORIGIN

215 a 143 c 171 g 195 t

Query Match 41.6%; Score 694; DB 10; Length 724;
Best Local Similarity 100.0%; Pred. No. 2.9e-179;
Matches 694; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ctgagaacccctcgaattcaatttcgcatcttgccagatgagaggaacaagtcttgct 74
Db 31 CTGAGAAACCCTCGAATTCATTTTCGCAATTTGGCAGAGATGAGAGGAACAAGTTTGTCT 90

QY 75 ttgatttacggtacactctctgcgtgctgctctgcgttcattacacagatgcgc 134
Db 91 TTGATTTACGGTACCTTCTCGCTGCTGCTCTCGCTTCATCTACATACAGATGCGGC 150

QY 135 ttctcgacacagtcagaatatgtagaccgcttgcgtgcgaattgaagcagaaaaatc 194
Db 151 TTTTCGGCACACAGTCAGATATGATGACCGCTTGTCTGCAATTTAAGCAGAAAATC 210

QY 195 attgtacaagtcagaccagattgcttattgacaagattaccagcagcaaggaagtag 254
Db 211 ATTGTACAAGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGGAAGACTAG 270

QY 255 tagctcttgaagaacaataaagcatcagaccagagtgccggcaattaaaggctcttg 314
Db 271 TAGCTCTTGAAGAACAATAATGAAGCATCAGGACCAGGAGTGCCGGCAATTAAGGGCTCTTG 330

QY 315 ttccagatcttgaagtaaggccataaaaaagtttaacgagagtgatgcagatgccagtgg 374
Db 331 TTCAGGATCTTGAAGTAAGGGCATAAAAAAGTTAATCGGAGATGTCAGATGCCAGTGG 390

QY 375 cagctgtagttgttatgcttgcgtcagctgactgactaccctgcagagagactattaaatcca 434
Db 391 CAGCTGTAGTGTATTGGCTTGCAGTCTGACTACCTTGCAGAGGACTATTAAATCCA 450

Qy	435	tcttaaaatacccaaacatctgtgtcatcaaaatcatcctcttttcatatccccagatgat	494
Db	451	TCCTAAAATACCAACAATCTCTGTCATCAAAATATCCCTCTTTTCATATCCCAAGGATGCAT	510
Qy	495	caaatcctgatgtaagaagaagtctgcttgaactatggtcagctgacgtatgcagcaact	554
Db	511	CAATCCCGATGTGAAGAAGCTGCTTTGAGCTATGTTGACGTACGCTACGCTATATGACAGCACT	570
Qy	555	tggattgaacctgtgcatactgaaagaccaggggaactggttgcatctactacaagattg	614
Db	571	TGGATTATGAACCTCTGCATACCTAAAGACACAGGGGAACCTGGTTCATACATACAAGATTG	630
Qy	615	cagctcattacaagtggcgatggtacagctgctttcacaaagcataaattttagccggtgta	674
Db	631	CACGTCATTAACAAGTGGGCATTTGGATCATCAGCTGTTTTCACAAGCATAAATTTTAGCCGTGTTA	690
Qy	675	tcatactagaagatgatggaatttgcgtcta	708
Db	691	TCATACTAGAGATGATATGGAATTCGCTGCTGA	724
RESULT 3			
BG889872			
LOCUS	EST515723	cSTD Solanum tuberosum cDNA clone	EST 30-MAY-2001
DEFINITION	661 bp	mRNA	linear
ACCESSION	EST515723	cSTD Solanum tuberosum cDNA clone	EST 30-MAY-2001
VERSION	BG889872	mRNA sequence.	
KEYWORDS	EST.		
SOURCE	potato.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanum.		
AUTHORS	van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.		
TITLE	Generations of ESTs from dormant potato tubers		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Cathy Ronning		
	The Institute for Genomic Research		
	For clone info: please contact Research Genetics, Libraries		
	Division tel 1-800-711-6195, email cdna@resgen.com		
	Seq primer: M13P-R.		
FEATURES	Location/Qualifiers		
source	1. 661		
	/organism="Solanum tuberosum"		
	/cultivar="Kennebec"		
	/db_xref="taxon:4113"		
	/clone="cSTD15H10"		
	/clone_lib="cSTD"		
	/tissue_type="dormant tuber"		
	/dev_stage="one month post-harvest"		
	/lab_host="SOLR"		
	/note="Vector: pBluescript SK(-); Site1: EcoRI; Site2: XhoI; This library targets genes expressed in dormant tubers. This library was made from sections of dormant tuber, avoiding the buds and epidermis. Tubers were stored for one month post-harvest at 40C. The tuber was peeled, well away from the surface. Then it was chopped into 1-2 mm cubes and immediately frozen in liquid nitrogen. This library is noted as P4 in Tanksley lab notebooks."		
BASE COUNT	196 a	133 c	158 g
ORIGIN	174 t		
Query Match	38.3%	Score 638.6;	DB 10; Length 661;
Best Local	Similarity 99.4%	Pred No. 4.4e-164;	
Matches	641; Conservative	0; Mismatches 4;	Indels 0; Gaps 0;
Qy	15	ctgagaacctogaatttcgatttcgagatgagaggaaacagtttgcgt	74

/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"

/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

BASE COUNT 180 a 124 c 145 g 193 t
ORIGIN

Query Match 37.8%; Score 631.4; DB 10; Length 642;
Best Local Similarity 99.1%; Pred. No. 4.1e-162;
Matches 635; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 346 gtaaatcgagatgacagatgccagtcgagctgtagtgtattgcttgacgtcgtac 405
Db 2 GTTAATCGGAGATGTCAGATGCCAGTGGCAGCTGTAGTTTATGGCTTGCAGTCGTGC 61
QY 406 tgactactggagaggactataatccatcttataaaataccaaacatctgttgatcaaa 465
Db 62 TGACTACTGGAGGAGCTATTAAATCCATCTTAAATACCAACATCTGTGCATCAAA 121
QY 466 atattcttttccatctccaggaatgatacaatccctgatgtaagaagctgtgttgag 525
Db 122 ATATCTCTCTTTTCATATCCAGGATGGATCAAAATCCCTGATGTAAAGAAAGCTTGTCTGAG 181
QY 526 ctatggctcagtcgactatgacagcacttgattgattgaacctgtgatactaaagacc 585
Db 182 CTATGATCAGCTGACGTATATGACGACTTGGATTTTGAACCTGTGCATCTGAAAGACC 241
QY 586 aggggaactggttgactactacaaagtgcacgtcatcaaaagtggtgacgt 645
Db 242 AGGGAACTGGTTGCATCTACAAAGATGGACGCTCATTAAGTGGCATTGGATCAGCT 301
QY 646 gttccaaagcataattttagccgtgttatacactagaagatgatatgaaattgctgc 705
Db 302 GTTTCACAAGCATAAATTTAGCCGTGTATCATACTAGAAGATGATATGGAATTTGCTGC 361
QY 706 tgattttttgactattttgagctgagctactcttcttgacagagacaagtcgattat 765
Db 362 TGATTTTGTGACTATTATTTGAGCTGGAGCTACTCTTCTTGACAGACACAAGTCGATTAT 421
QY 766 ggtatttcttcttgaatgacaatggacaaagcagctgcgtccaaagatcctgatgctct 825
Db 422 GGCTATTCTCTTGGAAATGACAAATGGACAAAGGCAGTTCCTCCAGATCCTGATGCTCT 481
QY 826 ttaccgctcagacttttttctggtcttgatgagcttctcaaaatacaacttggtccga 885
Db 482 TTACCGCTCAGACTTTTCTGCTGTGGATGGATGCTTTCAAAATCAACTTGGTCCGA 541
QY 886 actatctccaaagtggccaaaggttactgggagctgagctgaagctgaaagaaatca 945
Db 542 ACTATCTTCAAGTGGCCAAAGCTTACTGGATGACTGGCTGAAGCTGAAGAAATCA 601
QY 946 cagagtcacaacttatttcgcccaagatttgcagaacgt 986
Db 602 CAGAGGTGCAAAATTTATTACCCAGAAAGTTTGCAAAACGT 642

RESULT 5
AI486771
LOCUS
DEFINITION
EST245093 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
AI486771
ACCESSION
VERSION
AI486771.1 GI:4382142
KEYWORDS
EST
SOURCE
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 643)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers
source
1..643
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED11D20"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and
directionally cloned cDNA in vector Lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 179 a 128 c 144 g 192 t
ORIGIN

Query Match 36.9%; Score 615.8; DB 9; Length 643;
Best Local Similarity 97.4%; Pred. No. 7.8e-158;
Matches 626; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY 346 gtaaatcgagatgacagatgccagtcgagctgtagtgtattgcttgacgtcgtac 405
Db 1 GTTAATCGGAAATGTCAGATGCCAGTGGCAGCTGTAGTTTATGGCTTGCAGTCGTTC 60
QY 406 tgactactggagaggactataatccatcttataaaataccaaacatctgttgatcaaa 465
Db 61 TGACTACTCGGAGAGACTATAAAATCCATCTTAAATACCAACATCTGTTCATCAAA 120
QY 466 atattcttcttccatctccagatgatacaatccctgatgtaagaagctgtgttgag 525
Db 121 ATATCTCTCTTTTCATATCCAGGATGGATCGAATCCCTGTATGTTAGAAAGCTTGTCTTGAAG 180
QY 526 ctatggctcagctgactatgacagcacttgattgaacctgtgcatactgaaagacc 585
Db 181 CTATGATCAACTGACGTATATGACGACTTGGATTTTGAACCTGTGCATCTGAAAGACC 240
QY 586 aggggaactggttgactactacaaagtgcacgtcatcaaaagtggtgacgtcgtac 645
Db 241 AGGGAACTGGTTCATCTACTACAAAGATTGACGCTCATTAAGTGGCATTGGATCAGCT 300
QY 646 gttccaaagcataaattttagccgtgttatacactagaagatgatatgaaattgctgc 705
Db 301 GTTTCACAAGCATAAATTTTAGCCGTGTATCATACTAGAAGATGATATGGAATTTGGCGC 360
QY 706 tgattttttgactattttgagctgagctactcttcttgacagagacaagtcgattat 765
Db 361 TGATTTTGTGACTACTTTTGAGCTGGAGCTACTCTTCTTGACAGACACAAGTCGATTAT 420
QY 766 ggtatttcttcttgaatgacaatggacaaagcagctgcgtccaaagatcctgatgctct 825
Db 421 GGCTATTTCATCTTGGAAATGACAAATGGACAAAGGCAGTTCCTCCAGATCCTTATGCTCT 480
QY 826 ttaccgctcagacttttttctggtcttgatgagcttctcaaaatacaacttggtccga 885
Db 481 TTACCGCTCAGACTTTTCTGCTGTGTGGATGGATGCTTTCAAAATCAACTTGGTCCGA 540

QY	886	actatctccaaagtgcgcaaaaggcttactcgggatgcactggaagctggaagaaatca	945
Db	541	ACTATCTCCAAAGTGCAGAGGCCTTACTGGATGACTGGCTCAAGAAATCA	600
QY	946	caagagtcgaacattattcaccagaagtttcgaaacgtac	988
Db	601	CAGAGTGCACAATTATTTCGCCCAAGATTTCAGAACGFTC	643
RESULT	6		
LOCUS	AW979500	673 bp	linear
DEFINITION	EST1310521 tomato root deficiency, Cornell University Lycopersicon		EST 18-MAY-2001
ACCESSION	AW979500		esculentum cDNA clone cLEW209 5', mRNA sequence.
VERSION	AW979500.1	G1:8171019	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	1 (bases 1 to 673)		
AUTHORS	van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F., Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J., and Tanksley,S.D.		
TITLE	Generation of ESTs from tomato nutrient-deficient roots		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.		
FEATURES	Location/Qualifiers		
source	1..673		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA496"		
	/db_xref="taxon:4081"		
	/clone="cLEW209"		
	/clone_lib="tomato root deficiency, Cornell University"		
	/tissue_type="roots"		
	/dev_stage="5-6 weeks old"		
	/notes="vector: pBluescript SK-; Site_1: 5' EcoRI; Site_2: 3' XhoI; supplier: Tanksley; Tissue supplied by Dave Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested from plants grown under the following deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and mRNA was isolated from individual treatments. Proportional aliquots of mRNA of each treatment were mixed and used for library construction."		
BASE COUNT	198 a 136 c 159 g 180 t		
ORIGIN			
Query Match	35.9%;	Score 598.6;	DB 9;
Best Local Similarity	96.2%;	Pred. No. 4.2e-153;	
Matches	613;	Conservative	0; Mismatches 24; Indels 0; Gaps 0;
QY	20	aaacccctcgaattcaatttcgcatctggcagagatgagagggaacagtttcttgat	79
Db	37	AACCCTTAGAAATTCATTTCGCATTGGCAGAGATGAGAGGAACAAGTTTTCCTTTGAT	96
QY	80	ttacggttaccttctcgtcgtggtgctcctcgcttcattcatcacagatgcggcttttc	139
Db	97	TTACGGTACCTTCTCGTCGTGGTGTCTCGCCTTCATCTACATACAGATGCGGCTTTC	156
QY	140	gcacacagtcacaatatgtagaccgccttgcctgcattgaattgaacagaaatcatgtt	199
Db	157	GTGACACAGTCAGAAATGACAGACCGCTTGCCTGCGATTGACGAGAAATCATTTGT	216


```

BG131020
LOCUS      BG131020        602 bp      mRNA      linear      EST 31-JAN-2001
DEFINITION EST463912 tomato crown gall Lycopersicon esculentum cDNA clone
            cTOE2M21 5' sequence similar to Solanum tuberosum beta-1
            ,2-N-acetylglucosaminyl transferase 1, mRNA sequence.
ACCESSION  BG131020
VERSION    BG131020.1 GI:12631208
KEYWORDS   EST.
SOURCE     tomato.
ORGANISM   Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE  1 (bases 1 to 602)
AUTHORS   van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Hansen,C., Ronning
            ,C. and Tanksley,S.
TITLE     Generation of ESTs from tomato crown gall tissue
JOURNAL   Unpublished (2001)
COMMENT   Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES   Location/Qualifiers
            source
            1..602
            /organism="Lycopersicon esculentum"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cTOE2M21"
            /clone_lib="tomato crown gall"
            /tissue_type="crown gall"
            /dev_stage="crown galls from full-grown plants (8 wks old
            )"
            /lab_host="SOLR"
            /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; Four wk old greenhouse plants were stab inoculated
            on stem with Agrobacterium tumefaciens C58 (Dr. T.J. Burr,
            Cornell U.). Galls were allowed to develop for another 4
            wks, when gall tissue was frozen in liquid nitrogen."
BASE COUNT 193 a 103 c 137 g 169 t
ORIGIN
Query Match      28.7%; Score 479.6; DB 10; Length 602;
Best Local Similarity 89.7%; Pred. No. 1.8e-120;
Matches 540; Conservative 0; Mismatches 54; Indels 8; Gaps 2;

Qy 1061 gtccaggttgattgaagtcgaatgacctaaagttacaccttttggaggacaactatgtgaaa 1120
Db 1 GTGAAGGTTGATTGGAAGTCAATGCACCTAAGTTACCTTTTGGAGGACAACTATGTAAAA 60

Qy 1121 cactttggcgactgggttaaaaggctaagccccatccacggagctgatgctgttttgaaa 1180
Db 61 CACTTTGGAGACTTGGTTAAAGAGCGTTAAGCCCATCCATGCAGCTGATGCTGTTTGA 120

Qy 1181 gcatttaacatagatgctgattcagtcattcagtcagacagacacactagactttgaagat 1240
Db 121 GCATTTAACATAGATGCTGATGCTGCTGATTTCAGTACAGAGACCAACTAGACTTTGAAGAC 180

Qy 1241 atcgctcgacagtttggcattttttgaagaatggaagatgggttaccacggcgagcattat 1300
Db 181 ATCGCACGACAGTTTGGAAATTTTGAAGATGGAAGGATGGTGTACCACGGGACATAT 240

Qy 1301 aaagggatagtagtttccgggtttcaaacatctagacgtggttccctgtttccctgtat 1360
Db 241 AAAGGAATAGTACTTTTCCGGTTTTCAAACATCTAGACGTGTGTTCCCTTATTTGCCCTGAT 300

Qy 1361 tctcttcgacaacttgagttgaagatacttagcgaagatattgtagccctgagcaac 1420
Db 301 TCTCTTCAACAACCTTGGAGTTGAAGATACTTAGCAAGATATGATTGGCGGTGAGCAAC 360

Qy 1421 aattagactatttggtaggatacatttgaaagagctgacagaaagatgactacca 1480

```

```

Db 361 AATTTAGACTTATTGGCAGGATACATTTGAAGAGCTGCACGAAAGCTTGATTAGCA 420
Qy 1481 gtagctacatgcacattttaatgtaataaggagaaccactgcttattgttggatgg 1540
Db 421 GTAGCTGCATCAACAATTTTAATGTAAATGGAAGGAACCTACTGCCTACTGTGGAGTGG 480
Qy 1541 atgaatcatcaccac-----atcctattattcaagtttacaacaataagagaaaagttt 1595
Db 481 ATGAAATGTAGCCCCCATATAACACAATGTTTTCGTTACACAAAGAGGAGGAATGTA 540
Qy 1596 gccctataaaaaaaa-----tttttggttcttaagaaggaacgttaccgattatgaacactt 1652
Db 541 GCCCCCATATAACACAATCTTTTTCGTTCTTAAAGGAAGCACTTACGATTATACGCAACT 600
Qy 1653 tt 1654
Db 601 TT 602

RESULT 10
LOCUS   BG887316
DEFINITION EST513167 cSTD Solanum tuberosum cDNA clone CSTD4124 5' sequence,
            mRNA sequence.
ACCESSION BG887316
VERSION    BG887316.1 GI:14264402
KEYWORDS   EST.
SOURCE     potato.
ORGANISM   Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 575)
AUTHORS   van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE     Generations of ESTs from dormant potato tubers
JOURNAL   Unpublished (2001)
COMMENT   Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13P-R.
FEATURES   Location/Qualifiers
            1..575
            /organism="Solanum tuberosum"
            /cultivar="Kennebec"
            /db_xref="taxon:4113"
            /clone="cSTD4124"
            /clone_lib="cSTD"
            /tissue_type="dormant tuber"
            /dev_stage="one month post-harvest"
            /lab_host="SOLR"
            /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; This library targets genes expressed in dormant
            tubers. This library was made from sections of dormant
            tuber, avoiding the buds and epidermis. Tubers were stored
            for one month post-harvest at 40C. The tuber was peeled,
            well away from the surface. Then it was chopped into 1-2
            mm cubes and immediately frozen in liquid nitrogen. This
            library is noted as P4 in Tanksley lab notebooks."
BASE COUNT 180 a 98 c 120 g 177 t
ORIGIN
Query Match      26.5%; Score 441.6; DB 10; Length 575;
Best Local Similarity 97.9%; Pred. No. 4.7e-110;
Matches 458; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 1189 catagatggtgatgctgctattcagtcagacacactagactttgaagatcgctcg 1248
Db 2 CATAGATGGTGATGCTGCTATTTCAGTACAGAGACCAACTAGACTTTGAAGCTATCGCTCG 61

```

QY	1249	acagtttgccatcttttgaaagatgaaagatggtgtaccacggcgagcatataaaagguat	1308
Db	62	ACAGTTTGCCATTTTGAAGAATGAAGACGTGCTACACGGCGAGCATATAAAGGGAT	121
QY	1309	agtaatttcggtttcaaacatctagacgtgtgtctcttcttccctgattctcttcg	1368
Db	122	AGTAGTTTCCGGTTTCAAAACATCTAGACGCTGTGTCTTGTGCCCCCTGATTCCTTCG	181
QY	1369	acaacttgagttgaagatacttagcgaagataatgtaggagcctgagcaacaatttaga	1428
Db	182	ACAACCTGGAGTTGAAGTACTTACCGAAGATATGATTTGGAGCCTGAGCAACAATTAGA	241
QY	1429	cttatttgtagatcacatttgaaagagctgacacgaaagatgatgactaccagtagctac	1488
Db	242	CTTATTTGGTAGATACATTTGAAAGAGCTGACACGAAAGATGATGACTACCAGTAGCTAC	301
QY	1489	atgcaacattttaaattgaaggaaggaacccactgctattgttggaatggaatgaatca	1548
Db	302	ATGCACACATTTTAATGTTAATGAAGGAACCCACTGCTATTGTTGGAAATGGATGAATCA	361
QY	1549	tcaccacatcctattattcaag-tttcaaacataaagaggaaatgttgcctataaaaa	1607
Db	362	TCACCACATCCTATTATTCAAGTTTTCACAAACATAAAGAGGAAATGTAGCCCTATAAACA	421
QY	1608	caaatctttgtttctaagaaaggaacgttacgattatgagaacatttg	1655
Db	422	CAATGTTTTTGTCTTAAAGGAACGTTAGCATTTTGAGCAACTTTG	469
RESULT	11		
LOCUS	AI725875	681 bp mRNA linear	EST 11-JUN-1999
DEFINITION	BNLGH113374 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to ALPHA-1, 3-MANNOSYL-GLYCOPROTEIN, mRNA sequence.		
ACCESSION	AI725875		
VERSION	AI725875.1	GI:5044727	
KEYWORDS	EST		
SOURCE	upland cotton.		
ORGANISM	Gossypium hirsutum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.		
AUTHORS	1 (bases 1 to 681)		
TITLE	Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.		
JOURNAL	ESTs from developing cotton fiber		
COMMENT	Unpublished (1999) Contact: Ben Burr Biology Department Brookhaven National Laboratory Upton, NY 11973, USA Tel: 516-344-3396 Fax: 516-344-3407 Email: burr@bnl.bnl.gov		
FEATURES	Seq primer: T3 Primer. Location/Qualifiers		
source	1. 681 /organism="Gossypium hirsutum" /cultivar="Acala Maxxa" /db_xref="taxon:3635" /clone_lib="Six-day Cotton fiber" /tissue_type="immature fiber" /dev_stage="six days post anthesis" /lab_host="Xl1-Blue" /note="vector: pBluescript II KS+"		
BASE COUNT	202 a 129 c 151 g 197 t	2 others	
ORIGIN			
Query Match	25.9%	Score 431.6;	DB 9; Length 681;
Best Local Similarity	77.1%	Pred. No. 2.8e-107;	
Matches	524; Conservative 0;	Mismatches 136;	Indels 0; Gaps 0;
QY	506	gtagaagaagcttgcttgagctatggtcagctgacgtatgacgactatgacgacttgattatgaa	565

Db	1	GTAAAACTAAGGCTTTGAGTTTATAAGGAGCTAACTTATATGACGACATAGATTATGAT	60
QY	566	cctgtgcatactgaaagaccagggaactggtgtgcatactacaagatlgcacgtctattac	625
Db	61	CCGTTGCATACAGACGGGCTGGGAAATTGATCGCATACTACAAGATTGCCGCTCACTAC	120
QY	626	aagtgggcatgagtcagctgtttcacaaagcataattttagccgtgtatcatcatalagaa	685
Db	121	NAATGGCATTCGATGAGTTGTCTACAAGACCAATTTTGGACCGAGCTAATAATACITGAA	180
QY	686	gatgatatggaatctgctgctgatttttttgactatttttgaggtcggagctactctctt	745
Db	181	GATGATATGGAATTTGCCCTGATTTTGTGATTACTTTGAGGCAGCTCTGCCCTTCTC	240
QY	746	gacagagacaagtcgattatgctattctcttgaatgacaatggacaagagcgagtlc	805
Db	241	GACAAGGACAAGTCAATTTATGGCTGTTTCTCATGGAATGACAATGGCGCAAAAGCAGTTT	300
QY	806	gtccaagatcctgatgctctttaccgctcagacttttttccctgggtcttggatggatgctt	865
Db	301	GTGTATGACCCCATATGCACCTTTTATCGCTCAGATTTCTTTCTGGTCTTGGTGGATGCTT	360
QY	866	tcaaaatacaacttggccgaactatctccaaagtggccaaaggcttacttgggagatgactgg	925
Db	361	ACTAAATCTGTATGGAATGAGCTATCACAAAATGGCCAAAAGCTTACTTGGATGACTGG	420
QY	926	ctaaagctgaaagaaatcacagaggtcgacaatttattcggccagaaagtcttgcgaacg	985
Db	421	TTGAGATTAAAAAATAATCACAAATGGTCGACAAATTCCTTCGCTCGAAGTATGCAGAAC	480
QY	986	tacaattttggtgagcatggttctagtttggggcagttttttaaagcagtatctttagacca	1045
Db	481	TATAATTTTGGTGAGCATGGTTCAAGCATGGGCGAGTTTTTCGAAAAATACCTTGCACCT	540
QY	1046	attaagctaaatgatgtccaggttgatggaagtcaatgacctaagttacatttttggag	1105
Db	541	ATTAAGATGAATGACGTGNAGGTGGACTGGAAGTNCAGGATTTTGAGCTTACCTTAAC	600
QY	1106	gacaactatgaaacacttttggcagcttggtaaaaggctaaagccctccacgagct	1165
Db	601	GAATAATATGCCCACTACTTTCGACAGACATTTTCMAAGCGGCAAAACCTGTCCCTGGA	660
QY	1166	gatgctgttttgaagcatt	1185
Db	661	GATTCTGCTTTATGGCATT	680
RESULT	12		
LOCUS	AI897249	528 bp mRNA linear	EST 18-MAY-2001
DEFINITION	EST266608 tomato ovary, TAMU lycopersicon esculentum cDNA clone CLED26F21, mRNA sequence.		
ACCESSION	AI897249		
VERSION	AI897249.1	GI:5603067	
KEYWORDS	EST.		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.		
AUTHORS	1 (bases 1 to 528) Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.		
TITLE	Generation of ESTs from tomato carpel tissue		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: CUGI Clemson University Genomics Institute 100 Jordan Hall, Clemson, SC 29634, USA		

RESULT	14
AW618699	
LOCUS	530 bp mRNA linear EST 18-MAY-2001
DEFINITION	EST320685 L. pennellii trichome, Cornell University Lycopersicon pennellii cDNA clone cLPT14D3 5', mRNA sequence.
ACCESSION	AW618699
VERSION	
KEYWORDS	AW618699.1 GI:7324945
SOURCE	EST.
ORGANISM	Lycopersicon pennellii. Lycopersicon pennellii. Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
REFERENCE	1 (bases 1 to 530)
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Lakey,J., Holt,I.E., Liang,F., Hansen,T.S., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J. Generation of ESTs from wild tomato (<i>Lycopersicon pennellii</i>) trichomes
TITLE	
JOURNAL	Unpublished (1999)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html 5 prime sequence.
FEATURES	Location/Qualifiers
source	1..530
	/organism="Lycopersicon pennellii" /db_xref="taxon:28526" /clone="cLPT14D3" /clone_lib="f1.pennellii trichome, Cornell University" /tissue_type="trichome" /dev_stage="mixed stages" /lab_host="SOLR"
BASE COUNT	155 a 102 c 128 g 145 t
ORIGIN	
Query Match	23.7%; Score 395; DB 9; Length 530;
Best Local Similarity	96.4%; Pred. No. 2.8e-97;
Matches	404; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY	20 aaaccctcgaaattcaatttcgcatgtggcagatgcaggagaacaagtgttgattgat 79
Db	AACCCCTTAGAATTCAATTTTCGCAATTCGCAGAGTACAGGAACAAGTTTTGCTTTGAT 171
OY	80 ttacagttacctctcgtcgtcgtcgtcgtcccttcatacacagatcgccgttttc 139
Db	TTACGGTACCCTTCTCGTGCGGTCTCTCGCCCTTCATCTACATACAGATCGCGCTTTTC 231
OY	140 gcgacacagtcagaatatgtagaccgccttgtcgtcgaattgacgagaaaatcattgt 199
Db	GTGCACACAGTCAGAATATGCAGACGCCCTTCTGCTGCAATTGAAGCAGAAAATCATGTG 291
OY	200 acaagtgcagacagattgcttatigacaagattagccagcagcaagaagaagtagtagct 259
Db	ACAAGTCAGACTAGTGTGCTTATTGACAAGATTAGCCAGCAGCAAGGAAGTAGTGCT 351
OY	260 cttaaagaacaataaagactcaggaccagagtgccgcaaatgaaggctctgttcag 319
Db	CTTAAGAACAATAAAGCGCTCAGSACCAGGATGCCGACAAATTAGGGGCTCTTGTTTCAG 411
OY	320 gatcttgaaagtaaggcataaaaaagctaatacggagatgtgcagatgccagtggcagct 379

Db	62	AAAGCTTTGAGTTTACACTCAAAATAACATTTATGACGATGTGGATCTTGAGCCTGTGGCC	121
Qy	575	actgaagacacagggaactgggtgcatactacaagaatggcagctcattacaagtgga	634
Db	122	ACTCAAAAGCACGAGAGAAAACGTTGCATATTACAAGATAGCTAACCACTATATAATGGGCC	181
Qy	635	ttgatcagctggtttcacagaacaaatttagccggttatcatcatactagaagatgatg	694
Db	182	TTGATGAGCTATTTCATTAAAGCATGATTTTCGTCGAGTAATCATTTCTGGAAGATCACATG	241
Qy	695	gaaattgctgctgattttttgactattttgagctggagctactctcttgacagaagac	754
Db	242	GAGATCGCCCCAGATTTCTTTGACTACTTTGAGGCTGCGAGCGAAATTACTTCACACTGAC	301
Qy	755	aagtcgattatggctatbttctcttggaatgacaatggacaaagcagttcgctccaagat	814
Db	302	AACTCGAATAATGGCTGTTTCTTCTTGGAAATGACAATGGGCAAAACGAGTTGTTTATGAC	361
Qy	815	ccctgagctctttaccgcctcagaacttttctctgctcttgatgatgactttcaaaatca	874
Db	362	CCAAAGCTCTTTTACCCTGGGATTTCTTTCTCGGGCTTGGATGGATGCTTAACCAAGTCA	421
Qy	875	acttgctccgaactatctccaaagtgggccaaaggcttactgggatgactgctgaagctg	934
Db	422	ACATGGATGGAGCTGTACACAAAGTGGCCAAAGCTTATTCGGATCACTGGGTGAGGCTA	481
Qy	935	aaagaaaaatcacagagggtcgacaatttatccgccagaagtttcagacgtaacaatttt	994
Db	482	AAGGAGGTACAGAGATCGGCAGATTATTCGGCCAGAAAGTATGCAGAAACATACAACCTT	541
Qy	995	ggtgagcatggtctcagtttggggcagttttttaaagcagtatcttgagccaattaaagcta	1054
Db	542	GGCGACATGGATCAAGCATGGCACAATCTTTTGATCAACTACTTGAACCAATCAAAATTA	601
Qy	1055	aatgatgtccaggttggattggaagtc	1080
Db	602	AATGATGCTCATATTTCAGTGGAAATTC	627

Search completed: August 13, 2002, 19:34:14
Job time: 7282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:55:12 ; Search time 83.73 Seconds
(without alignments)
4896.242 Million cell updates/sec

Title: US-09-591-466C-1
Perfect score: 1669
Sequence: 1 gaattcgccgcctgaga.....actttggcgcgcgaattc 1669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	37.4	2.2	2934	1	US-08-198-446B-16
C 2	37.4	2.2	2934	2	US-08-870-693-16
C 3	37.2	2.2	7218	1	US-08-232-463-14
4	35.2	2.1	2939	4	US-09-276-531-80
5	35	2.1	1058	4	US-08-858-207A-127
6	34	2.0	1716	3	US-08-656-034-9
7	34	2.0	2160	3	US-08-656-034-1
8	33.8	2.0	3360	1	US-07-712-833A-1
9	33.6	2.0	9472	1	US-08-325-547-9
C 10	32.8	2.0	472	4	US-09-269-617-2
11	32.8	2.0	2441	1	US-08-920-812-12
12	32.8	2.0	2441	1	US-08-920-827-12
13	32.8	2.0	2441	1	US-08-921-177-12
14	32.8	2.0	2441	1	US-08-362-57C-12
15	32.8	2.0	2441	2	US-08-920-828-12
16	32.6	2.0	1183	1	US-08-188-582-25
17	32.6	2.0	1183	1	US-08-646-715-25
C 18	32.6	2.0	2801	3	US-08-747-221B-30
19	32.6	2.0	2801	3	US-08-747-221B-32
C 20	32.6	2.0	2801	4	US-09-005-051-30
21	32.6	2.0	2801	4	US-09-005-051-32
C 22	32.6	2.0	2836	3	US-08-747-221B-24
23	32.6	2.0	2836	3	US-08-747-221B-26
C 24	32.6	2.0	2836	4	US-09-005-051-24
25	32.6	2.0	2836	4	US-09-005-051-26
26	32.2	1.9	2265	3	US-09-369-618-3
27	32.2	1.9	2265	3	US-09-369-617-3

C 28	31.8	1.9	3479	3	US-08-714-918-7	Sequence 7, Appli
C 29	31.8	1.9	3479	4	US-09-265-315-7	Sequence 7, Appli
C 30	31.8	1.9	3479	4	US-09-265-315-7	Sequence 7, Appli
C 31	31.8	1.9	3479	4	US-09-266-417-7	Sequence 7, Appli
C 32	31.8	1.9	10968	2	US-08-680-327-2	Sequence 2, Appli
C 33	31.8	1.9	10968	4	US-09-228-246-1	Sequence 1, Appli
C 34	31.8	1.9	3116	1	US-08-149-103-2	Sequence 2, Appli
C 35	31.6	1.9	3116	1	US-08-451-883-2	Sequence 2, Appli
C 36	31.4	1.9	624	4	US-09-397-99A-6	Sequence 5, Appli
C 37	31.4	1.9	1868	2	US-08-960-022-5	Sequence 6, Appli
C 38	31.4	1.9	80246	4	US-09-078-294-4	Sequence 4, Appli
C 39	31.4	1.9	80595	4	US-09-078-294-3	Sequence 3, Appli
C 40	31.2	1.9	398	3	US-08-714-918-5	Sequence 4, Appli
C 41	31.2	1.9	398	4	US-09-265-315-5	Sequence 5, Appli
C 42	31.2	1.9	398	4	US-09-265-315-5	Sequence 5, Appli
C 43	31.2	1.9	398	4	US-09-266-417-5	Sequence 5, Appli
C 44	31	1.9	624	4	US-09-397-99A-3	Sequence 3, Appli
C 45	30.8	1.8	578	4	US-09-020-956-104	Sequence 104, App

ALIGNMENTS

RESULT 1
US-08-198-446B-16/c
; Sequence 16, Application US/08198446B
; Patent No. 5674996
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/198,446B
; FILING DATE: 18-FEB-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCRI7537
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: yeast MEC2 cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 395..2724
US-08-198-446B-16


```

Query Match          2.2%; Score 37.4; DB 1; Length 2934;
Best Local Similarity 48.4%; Pred. No. 0.16;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 621 attacaagtgccgcatggatcagctgtttcacagaacataattttagccgtgttatcatc 680
    || || || || || || || || || || || || || || || || || || || || ||
Db 2020 ATATCGAATAGGGCGGTGTCGGGTATTTTAAATCTTGAATTTTCCATCTTGGTCTTC 1961

QY 681 taagaatgatgaataatgctgctgatttttttgactattttgagcggtgagctactc 740
    || || || || || || || || || || || || || || || || || || || || ||
Db 1960 CTGATCCTGTTCTTGAAGTTGTTGCTCCATTTTCTTAATAATTTCTGCTCGCACACGATTTGCTAAGGA 1841

QY 741 ttcttgacagagacaagtcgatttgctatttcttcttgggaatgacaaatgggacaagcg 800
    || || || || || || || || || || || || || || || || || || || || ||
Db 1900 GTATTGAGCATCGTCCATATTTTCTTAATAATTTCTGCTCGCACACGATTTGCTAAGGA 1841

QY 801 agtctccaagatcctgatcgtctctttaccgctca 835
    || || || || || || || || || || || || || || || || || || || || ||
Db 1840 TATTGTGAAAAATCACCATATGATTGTGAGGCCA 1806

RESULT 2
US-08-870-693-16/c
; Sequence 16, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCRI10798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA

```

```

; DESCRIPTION: yeast MEC2 cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 395..2724
US-08-870-693-16

Query Match          2.2%; Score 37.4; DB 2; Length 2934;
Best Local Similarity 48.4%; Pred. No. 0.16;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 621 attacaagtgccgcatggatcagctgtttcacagaacataattttagccgtgttatcatc 680
    || || || || || || || || || || || || || || || || || || || || ||
Db 2020 ATATCGAATAGGGCGGTGTCGGGTATTTTAAATCTTGAATTTTCCATCTTGGTCTTC 1961

QY 681 tagaagatgatattggaattgctgctgatttttttgactattttgagcggtgagctactc 740
    || || || || || || || || || || || || || || || || || || || || ||
Db 1960 CTGATCCTGTTCTTGAAGTTGTTGCTCCATTTTCTTAATAATTTCTGCTCGCACACGATTTGCTAAGGA 1841

QY 741 ttcttgacagagacaagtcgatttgctatttcttcttgggaatgacaaatgggacaagcg 800
    || || || || || || || || || || || || || || || || || || || || ||
Db 1900 GTATTGAGCATCGTCCATATTTTCTTAATAATTTCTGCTCGCACACGATTTGCTAAGGA 1841

QY 801 agtctccaagatcctgatcgtctctttaccgctca 835
    || || || || || || || || || || || || || || || || || || || || ||
Db 1840 TATTGTGAAAAATCACCATATGATTGTGAGGCCA 1806

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```


; TELEFAX: 610-270-5090

; TELEX:

; INFORMATION FOR SEQ ID NO: 127:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1058 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-858-207A-127

Query Match

Best Local Similarity 2.1%; Score 35; DB 4; Length 1058;

Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 454 ttttcataaaatctcttttccatccccagatggtacaaatcctgatgtaagaaa 513

DB 640 TGGTCAGAAAATTATCTCCTTACTTACGAAAAGAACGCTGAGGATACAGAGAAATCTT 699

QY 514 gcttgcttgagctatggtcagctgacgtatgacgacctggattatgaacctgtgca 573

DB 700 GCTTCGTTAGCCACTATTATCTCGAGCAGGAGCGTTATGAGGATATTTCTAGACTTGCA 759

QY 574 tactgaagaccaggggaactggtgtgcatactactacaagattgcacgtcatta 624

DB 760 GAGTCAGGAGCCAGAAAATCTTTTACCAAGTGGATGATGCTGCTCTTTA 810

RESULT 6

US-08-656-034-9

; Sequence 9, Application US/08656034

; Patent No. 6015691

; Patent No. 6015691 5786172

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED

; TITLE OF INVENTION: ADHESION PROTEIN GENES OF EHRlichia

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/656.034

; FILING DATE: Concurrently Herewith

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Corder, Timothy S.

; REGISTRATION NUMBER: 38,414

; REFERENCE/DOCKET NUMBER: UTSG:191

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (512) 474-7577

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1716 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 57..1700

US-08-656-034-9

Query Match

Best Local Similarity 2.0%; Score 34; DB 3; Length 1716;

Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1493 aacattttaattgaaggagaccactcttattgttggaatggatgaatcac 1552

DB 1219 AAGTTTCTAAAGTTGAACAAGAAAAAACAACCTGAAATTTCTAGTAGAAGATTTGCCAT 1278

QY 1553 cacatcctctattattcaagtttacaaacataaagagagaaatgttgccctataaaaaa 1612

DB 1279 TAGGTCAAGTGATTCGGGTGTGTGTAGACAAGATGTTTGACCTTCATTATATC 1338

QY 1613 ttttggtttctaagaagaacgttaacgattatga 1646

DB 1339 CAATCGTTATAAAGGAGGAAGATAAAGTTTGTGA 1372

RESULT 7

US-08-656-034-1

; Sequence 1, Application US/08656034

; Patent No. 6015691

; Patent No. 6015691 5786172

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: IMMUNODOMINANT 120 KDA SURFACE-EXPOSED

; TITLE OF INVENTION: ADHESION PROTEIN GENES OF EHRlichia

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Arnold, White & Durkee

; STREET: P.O. Box 4433

; CITY: Houston

; STATE: Texas

; COUNTRY: United States of America

; ZIP: 77210

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/656.034

; FILING DATE: Concurrently Herewith

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Corder, Timothy S.

; REGISTRATION NUMBER: 38,414

; REFERENCE/DOCKET NUMBER: UTSG:191

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (512) 418-3000

; TELEFAX: (512) 474-7577

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2160 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 171..2054

US-08-656-034-1

Query Match

Best Local Similarity 2.0%; Score 34; DB 3; Length 2160;

Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 1493 aacattttaattgaaggagaccactcttattgttggaatggatgaatcac 1552

DB 1573 AAGTTTCTAAAGTTGAACAAGAAAAAACAACCTGAAATTTCTAGTAGAAGATTTGCCAT 1632


```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORGANISM: Enterococcus faecalis
; STRAIN: Clinical Isolate S2-27
; US-08-920-827-12

Query Match                2.0%; Score 32.8; DB 1; Length 2441;
Best Local Similarity      52.1%; Pred. No. 4.1;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 743 ctgtacagacagtcgattggtcttcttccgctcagactttttccctggatggaagcgag 802
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1848 CTTTCCTAAGATAACTAGAAATTTTCTTACGCTCTCAGAAAGCCAAAGCTCAATTATTG 1907

QY 803 ttgtccaagatcgtgatgctctttaccgctcagactttttccctggatggaagcgag 862
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1908 TGATTACCTTATAATCTCTCTCTTTATTCGGCGACCTCTTTAATATGATTATTGGAGG 1967

QY 863 ctttcaaaatcaactggctc 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1968 TTTTAAATTTGAAAGCTGTC 1987

RESULT 13
US-08-921-177-12
; Sequence 12, Application US/08921177
; Patent No. 5798211
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/08/921,177
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORGANISM: Enterococcus faecalis
; STRAIN: Clinical Isolate S2-27
; US-08-921-177-12

Query Match                2.0%; Score 32.8; DB 1; Length 2441;
Best Local Similarity      52.1%; Pred. No. 4.1;
Matches 73; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 743 ctgtacagacagtcgattggtcttcttccgctcagactttttccctggatggaagcgag 802
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1848 CTTTCCTAAGATAACTAGAAATTTTCTTACGCTCTCAGAAAGCCAAAGCTCAATTATTG 1907

QY 803 ttgtccaagatcgtgatgctctttaccgctcagactttttccctggatggaagcgag 862
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1908 TGATTACCTTATAATCTCTCTCTTTATTCGGCGACCTCTTTAATATGATTATTGGAGG 1967

QY 863 ctttcaaaatcaactggctc 882
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1968 TTTTAAATTTGAAAGCTGTC 1987

RESULT 14
US-08-362-577C-12
; Sequence 12, Application US/08362577C
; Patent No. 5807673
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,577C
; FILING DATE: 27-MAR-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
```

Query Match 2.0%; Score 32.8; DB 2; Length 2441;
Best Local Similarity 52.1%; Pred. No. 4.1;

[illegible]

RESULT 15
US-08-920-828-12
: Sequence 12, Application US/08920828
: Patent No. 5853998
: GENERAL INFORMATION:
: APPLICANT: Ohno, Tsuneya
: APPLICANT: Matsuhisa, Akio
: APPLICANT: Uehara, Hirotsugu
: APPLICANT: Eda, Soji
: TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
: NUMBER OF SEQUENCES: 25
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/920,828
: FILING DATE: 29-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/362,577
: FILING DATE: 27-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Rin-Laures, Li-Hsien
: REGISTRATION NUMBER: 33,547
: REFERENCE/DOCKET NUMBER: 19036/32420
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 18:14:32 ; Search time 285.38 Seconds
(without alignments)
10041.104 Million cell updates/sec

Title: US-09-591-466c-1
Perfect score: 1669
Sequence: 1 gaattgcgcgcgcctgaga.....actttgcgcgcgcgaattc 1669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802: *
1: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1669	100.0	1669	20 AAX78001	Potato GNTI cDNA.
2	1255	75.2	1737	20 AAX78002	Tobacco GNTI cDNA.
3	751	45.0	1639	21 AAC39540	Arabidopsis thaliana
4	728.6	43.7	1641	21 AAC50197	Arabidopsis thaliana
5	329.2	19.7	510	20 AAX78003	A. thaliana GNTI-h
6	192.8	11.6	2557	15 AAQ62625	Rat N-acetyl-gluco
7	181.6	10.9	2485	13 AAQ25594	Rabbit Gnt I cDNA
8	165.6	9.9	1335	22 AA169837	Human beta-1,2-N-a
9	164	9.8	3230	23 AAQ25595	Human Gnt I cDNA C

10	144.2	8.6	2250	22 AAF93806	Human cDNA encodin
11	131.4	7.9	2367	23 ABL22693	Drosophila melanog
12	121.2	7.3	335	21 AAC69652	Human acetylglucos
13	58.6	3.5	4721	23 ABL22692	Drosophila melanog
14	44	2.6	2078	22 AAH15975	Human cDNA sequenc
15	44	2.6	2353	22 AAH158317	Human polynucleoti
16	44	2.6	2372	21 AAF18300	Lung cancer associ
17	44	2.6	2711	22 AA160103	Human polynucleoti
18	44	2.6	2725	21 AA337048	Human polynucleoti
19	44	2.6	2725	22 AAS21483	Human cDNA sequenc
20	44	2.6	2725	22 AAD02922	Human cDNA sequenc
21	44	2.6	2725	22 AAF54262	DNA encoding prote
22	44	2.6	2725	22 AAF54262	Human cDNA sequenc
23	44	2.6	2731	21 AAZ94210	Human Prol1475 cDNA
24	41.2	2.5	725	22 AAF94005	Human transferase
25	40.2	2.4	4590	22 AAH24065	Primer specific fo
26	39.8	2.4	300	21 AAA01130	Yeast AOD9604-asso
27	39	2.3	11260	22 AAS45315	Human colon cancer
28	38.8	2.3	11473	20 AAX34650	Chemically pretrea
29	38.8	2.3	11475	22 AAH78338	Starch branching e
30	38.4	2.3	343	21 AAX42092	Nucleotide sequenc
31	38	2.3	910715	20 AAX20248	Human secreted exp
32	37.8	2.3	14041	22 AAH48024	Borrelia burgdorfe
33	37.4	2.2	2934	18 AAT91040	Internal control B
34	37.4	2.2	2934	20 AAX01275	Yeast checkpoint c
35	37.4	2.2	3299	22 AAH54896	Yeast MEC2 coding
36	37.4	2.2	4900	22 AAD05670	S. epidermidis gen
37	37.4	2.2	12842	22 ABA20815	Staphylococcus epi
38	37.2	2.2	14316	22 AAS61445	Human nervous syst
39	37	2.2	2328	20 AAX99549	Human gene regulat
40	36.6	2.2	2265	22 AAS40042	Nucleic acid sequ
41	36.6	2.2	2265	22 AAK91459	Genomic sequence #
42	36.4	2.2	5059	20 AAX84332	Human digestive sy
43	36.2	2.2	16768	23 ABL15650	Stealth virus nucl
44	36	2.2	400	22 AA185650	Drosophila melanog
45	36	2.2	7885	20 AAX76578	Human polynucleoti
					Human pancreas-spe

ALIGNMENTS

RESULT 1
AAX78001
ID AAX78001 standard; cDNA; 1669 BP.
AC AAX78001;
DT 19-AUG-1999 (first entry)
XX Potato GNTI cDNA.
DE
KW GNTI: beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;
KW deficient; defective; detection; transgenic plant; sugar residue;
KW medicine; antigenicity; deglycosylation; potato; ds.
OS Solanum tuberosum.
FH Key Location/Qualifiers
CDS 53..1394
FT /*tag= a
FT /product= "GNTI"
XX
XX DE19754622-A1.
XX
XX PD 10-JUN-1999.
XX
XX PF 09-DEC-1997; 97DE-1054622.
XX
XX PR 09-DEC-1997; 97DE-1054622.
XX
XX PA (VSCH/) VON SCHAEWEN A.
XX Von Schaewen A;
PI

QY	1192	agatggtgatgtcgtattcagttacagagaccactagactttgaagatatcgtcgcaca	1251	PR	04-MAY-1999;	99US-0132484.
Db	1266	agatggtgatgtcgtattcagttacagagaccactagactttgaaatatcgcacgcca	1325	PR	05-MAY-1999;	99US-0132485.
QY	1252	gtttggtcatttttgaagaatggaagatggtgtaccacgggcagcatataaauggatagt	1311	PR	06-MAY-1999;	99US-0132486.
Db	1326	atttggcatttttgaagaaatggaagatggtgtaccacgggcagcatataaauggatagt	1385	PR	07-MAY-1999;	99US-0132487.
QY	1312	agttttccggtttccaacatctagacgtgtgttccctgtttccctgattctcttcgcaca	1371	PR	11-MAY-1999;	99US-0132863.
Db	1386	agttttccggtttccaacatctagacgtgtgttccctgtttccctgattctcttcgcaca	1445	PR	14-MAY-1999;	99US-0134256.
QY	1372	acttgaagtgaagatacttagcgaagatatgatt---ggagccttgagacaacaatttga	1428	PR	14-MAY-1999;	99US-0134218.
Db	1446	actcgaattgaagatacttaacaagatatgattgcagagcccgccgcaaaatttttga	1505	PR	14-MAY-1999;	99US-0134219.
QY	1429	cttatttggtagatcacatttgaagagctgcacacgaagaagttagtactaccagtagctac	1488	PR	14-MAY-1999;	99US-0134221.
Db	1506	cttatttggtagatgcacgtgcacactaaacccatgattt-----accagttac	1558	PR	18-MAY-1999;	99US-0134370.
QY	1489	atcacaatttttaatttgaatggaagaagaccactgct-----t	1527	PR	18-MAY-1999;	99US-0134768.
Db	1559	atcacaatttttaatttgaatggaagaagaccactgct-----t	1618	PR	20-MAY-1999;	99US-0134941.
QY	1528	atttgggaatggtgaatcattcaccacacacatcctattat---ttcaagtttacaacataaag	1585	PR	21-MAY-1999;	99US-0135124.
Db	1619	ttcttagtattgtagaatacattcacaacacacacattatttttaagtgttcagaacataaag	1678	PR	24-MAY-1999;	99US-0135353.
QY	1586	aggaatgttgcctctataaaaacaa	1610	PR	25-MAY-1999;	99US-0136021.
Db	1679	aggaatgttgcctctataaaaacaa	1703	PR	27-MAY-1999;	99US-0136392.
RESULT	3			PR	28-MAY-1999;	99US-0136782.
AAC39540				PR	01-JUN-1999;	99US-0137222.
XX				PR	03-JUN-1999;	99US-0137528.
AC	AAC39540;			PR	04-JUN-1999;	99US-0137502.
DT	17-OCT-2000 (first entry)			PR	07-JUN-1999;	99US-0137724.
XX				PR	08-JUN-1999;	99US-0138094.
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 25006.			PR	10-JUN-1999;	99US-013847.
KW	Hybridisation assay; genetic mapping; gene expression control;			PR	14-JUN-1999;	99US-013847.
KW	protein identification; signal transduction pathway;			PR	16-JUN-1999;	99US-0139452.
KW	metabolic pathway; promoter; termination sequence; ss.			PR	17-JUN-1999;	99US-0139452.
XX				PR	18-JUN-1999;	99US-0139452.
OS	Arabidopsis thaliana.			PR	18-JUN-1999;	99US-0139457.
PN				PR	18-JUN-1999;	99US-0139458.
XX				PR	18-JUN-1999;	99US-0139459.
XX				PR	18-JUN-1999;	99US-0139460.
XX				PR	18-JUN-1999;	99US-0139461.
XX				PR	18-JUN-1999;	99US-0139462.
XX				PR	18-JUN-1999;	99US-0139463.
XX				PR	18-JUN-1999;	99US-0139750.
XX				PR	21-JUN-1999;	99US-0139763.
XX				PR	22-JUN-1999;	99US-0139817.
XX				PR	23-JUN-1999;	99US-0139899.
XX				PR	23-JUN-1999;	99US-0140353.
XX				PR	24-JUN-1999;	99US-0140354.
XX				PR	28-JUN-1999;	99US-0140695.
XX				PR	29-JUN-1999;	99US-0140823.
XX				PR	30-JUN-1999;	99US-0140991.
XX				PR	01-JUL-1999;	99US-0141287.
XX				PR	02-JUL-1999;	99US-0141842.
XX				PR	06-JUL-1999;	99US-0142055.
XX				PR	08-JUL-1999;	99US-0142390.
XX				PR	09-JUL-1999;	99US-0142803.
XX				PR	12-JUL-1999;	99US-0142920.
XX				PR	13-JUL-1999;	99US-0142977.
XX				PR	14-JUL-1999;	99US-0143542.
XX				PR	15-JUL-1999;	99US-0143624.
XX				PR	16-JUL-1999;	99US-0144005.
XX				PR	16-JUL-1999;	99US-0144085.
XX				PR	19-JUL-1999;	99US-0144086.
XX				PR	19-JUL-1999;	99US-0144331.
XX				PR	19-JUL-1999;	99US-0144332.
XX				PR	19-JUL-1999;	99US-0144333.
XX				PR	19-JUL-1999;	99US-0144334.
XX				PR	19-JUL-1999;	99US-0144335.
XX				PR	20-JUL-1999;	99US-0144352.
XX				PR	20-JUL-1999;	99US-0144632.
XX				PR	21-JUL-1999;	99US-0144884.
XX				PR	21-JUL-1999;	99US-0144814.
XX				PR	21-JUL-1999;	99US-0145086.
XX				PR	21-JUL-1999;	99US-0145088.

QY	845	ctggcttggatggatggttttcaaaatcaacttggtcgaactatctccaagtgcca	904	PR	16-APR-1999;	99US-0129845.
Db	818	ctggcttggatggatgctcaagatcgacttggatgagtatcaccaaagtgcca	877	PR	19-APR-1999;	99US-0130077.
QY	905	aaggcttactggatgactggttaaggctgaagaaaaatcacagaggtgcacaaattatt	964	PR	21-APR-1999;	99US-0130449.
Db	878	aaggcttactggatgactggttgagactaaaggaaaaaccataaagccgcacattcatt	937	PR	23-APR-1999;	99US-0130510.
QY	965	cgcccaagaagtttgcagaacgctacaattttggtgagcagtggtctagtttggggcagttt	1024	PR	28-APR-1999;	99US-0130891.
Db	938	cgaccggaagtcgtlagaacatacaaatittggtgaactgggtctagtttgggacagttt	997	PR	30-APR-1999;	99US-0131449.
QY	1025	tttaagcagatcttgagcccaattaaagctaaatgatgtccagggttgattggaagcaatg	1084	PR	PR	99US-0132048.
Db	998	ttcagtcagatcttggaacctataaagctaaacgatgtgacggttgactggaaagcaaa	1057	PR	PR	99US-0132407.
QY	1085	gacctaaagttaaccttttggagcaactatgtgaacactttggcgaacttggttaaaaag	1144	PR	PR	99US-0132484.
Db	1058	gacctgggataacctgacagagggaacctataccaaagtactttctggttagtgagaaa	1117	PR	PR	99US-0132485.
QY	1145	gctaagcccatccacggagctgctgtgttttgaagcatttaacatagatggtgatgtg	1204	PR	PR	99US-0132486.
Db	1118	gcacgaccaattcgaaggttctgaccttgtcttaaggctcaaaacataaaagatgatgtt	1177	PR	PR	99US-0132487.
QY	1205	cgtattcagtcagagacccaactagactttgaagatatacgtcgcagactttggcagttt	1264	PR	PR	99US-0132863.
Db	1178	cgtatccggtataaagacccaagtagagtttgaacgcatctgcagcagtttggcagttt	1237	PR	PR	99US-0134256.
QY	1265	gaagaatggaaggatggtgtaccacggcgagcatataaaggatagtagtttccggttt	1324	PR	PR	99US-0134219.
Db	1238	gaagaatggaaggatggtgtccctcgacacagcatataaaggatgtaggtttcgaatc	1297	PR	PR	99US-0134221.
QY	1325	caaacatctagacgtgttccctgttttccctgattctctcgacaacttggagttgaa	1384	PR	PR	99US-0134370.
Db	1298	cagacaacagacgtgttccctgttttccctgattctctcgacaacttggagttgaa	1357	PR	PR	99US-0134768.
QY	1385	gatacttagcgaaga	1399	PR	PR	99US-0134941.
Db	1358	aattcctgaigcaaa	1372	PR	PR	99US-0135124.
RESULT	4			PR	PR	99US-0135353.
AAC50197				PR	PR	99US-0135629.
XX	AAC50197 standard; DNA; 1641 BP.			PR	PR	99US-0136021.
AC				PR	PR	99US-0136392.
XX				PR	PR	99US-0136782.
DT				PR	PR	99US-0137222.
XX				PR	PR	99US-0137528.
DE	18-OCT-2000 (first entry)			PR	PR	99US-0137502.
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 63936.			PR	PR	99US-0137724.
XX	Hybridisation assay; genetic mapping; gene expression control;			PR	PR	99US-0138094.
KW	protein identification; signal transduction pathway;			PR	PR	99US-0138540.
KW	metabolic pathway; promoter; termination sequence; ss.			PR	PR	99US-0138847.
XX	Arabidopsis thaliana.			PR	PR	99US-0139119.
XX	EP1033405-A2.			PR	PR	99US-0139452.
PN				PR	PR	99US-0139453.
XX				PR	PR	99US-0139454.
PD				PR	PR	99US-0139455.
XX	06-SEP-2000.			PR	PR	99US-0139456.
PF				PR	PR	99US-0139457.
XX	25-FEB-2000; 2000EP-0301439.			PR	PR	99US-0139458.
PR				PR	PR	99US-0139459.
PR	25-FEB-1999; 99US-0121825.			PR	PR	99US-0139460.
PR	05-MAR-1999; 99US-0123180.			PR	PR	99US-0139461.
PR	09-MAR-1999; 99US-0123548.			PR	PR	99US-0139462.
PR	23-MAR-1999; 99US-0125788.			PR	PR	99US-0139463.
PR	25-MAR-1999; 99US-0126264.			PR	PR	99US-0139750.
PR	29-MAR-1999; 99US-0126785.			PR	PR	99US-0139763.
PR	01-APR-1999; 99US-0127462.			PR	PR	99US-0139817.
PR	06-APR-1999; 99US-0128234.			PR	PR	99US-0139899.
PR	08-APR-1999; 99US-0128714.			PR	PR	99US-0140353.
				PR	PR	99US-0140354.
				PR	PR	99US-0140695.
				PR	PR	99US-0140823.
				PR	PR	99US-0140991.
				PR	PR	99US-0141287.
				PR	PR	99US-0141842.
				PR	PR	99US-0142154.
				PR	PR	99US-0142055.
				PR	PR	99US-0142390.
				PR	PR	99US-0142803.
				PR	PR	99US-0142920.
				PR	PR	99US-0142977.
				PR	PR	99US-0143542.
				PR	PR	99US-0143624.
				PR	PR	99US-0144005.
				PR	PR	99US-0144085.
				PR	PR	99US-0144086.
				PR	PR	99US-0144325.
				PR	PR	99US-0144331.
				PR	PR	99US-0144332.
				PR	PR	99US-0144333.


```
QY 722 tttagagctgagactactcttctttgacagagacaaagtcgattatggtctatttcttcttgg 781
D 696 tttagagctgagactactcttctttgacagagacaaagtcgattatggtctatttcttcttgg 755
QY 782 aatgacaatggacaaaggcaggttcgtccagatccctgagctctttaccgctcagacttt 841
D 756 aatgataatggacaaaggcaggtttgtcgtatccctatgctctaccgacagacttt 815
QY 842 ttctctggtcttgatggtcttcaaaatcaacttggctccgaactatctccaaagtgg 901
D 816 ttctctggtcttgatggtctcaagagatcgacttggatgagttatccacaaagtgg 875
QY 902 ccaaaagg-cttactggatgactggttaaggtgaaagaaatcacagaggtcgacaatt 960
D 876 ccaaaaggcttactggatggttggtgagactaaaggaaacataaaagccgcaatt 935
QY 961 tattcgcacaaagtttgcagaacgltacaattttggtgagcaggttcttagtttgggca 1020
D 936 cattcgaccggaagtctgtagaacatacaattttggtgaacatggcttagtttgggca 995
QY 1021 gtttttaagcagatctcttgagccaattaaagctaaatgatgtccaggttgattggaagtc 1080
D 996 gtttttcagtcagatctctggaacctataaagctaaacgattgacggttgactggaaagc 1055
QY 1081 aatggacctaaagtactcttttgagagacaaactatgtgaacacttttggtgacttggtaa 1140
D 1056 aaaggacctgggtatcctgacagagggaaactataccaagtagctttcttggttagtgag 1115
QY 1141 aaaggctaaagccatccacggagctgctgttttgaagcaatttaacatagatggtga 1200
D 1116 acaagcacgaccaattcaagttctgacctgtcttaaggctcaaaacataaagatga 1175
QY 1201 tgtcgatttcagacagagaccactagactttgaaagatatgcctgcagtttggcat 1260
D 1176 tgtctgattccgtataaagaccaaagtaggtttgaacgattgcaagggaatttggtat 1235
QY 1261 ttctgaagaatgaagatggtgtaccacggcagcagcatataaaggatagtagtttccg 1320
D 1236 atttgaagaatgaagatggtgtgctcgaacagcatataaaggagtagtgggttccg 1295
QY 1321 gtttcaaacatctagacgtgtgttctctgttttccctgattcttctcgacaaacttgagt 1380
D 1296 aatccagacaaacagcgtgtattctctgttggccagatctctgtaatgcagcttggat 1355
QY 1381 tgaagatacttagcgaaga 1399
D 1356 tcgaattctctgatgcaaa 1374

RESULT 5
AAX78003
ID AAX78003 standard; cDNA; 510 BP.
XX
AC AAX78003;
XX
DT 19-AUG-1999 (first entry)
DE
DE A. thaliana GnTI-homologue cDNA fragment.
XX
KW GnTI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;
KW deficient; defective; detection; transgenic plant; sugar residue;
KW medicine; antigenicity; deglycosylation; ds.
XX
OS Arabidopsis thaliana.
XX
PN DE19754622-A1.
XX
PD 10-JUN-1999.
XX
PF 09-DEC-1997; 97DE-1054622.
XX
PR 09-DEC-1997; 97DE-1054622.
```

```
XX (VSCH/) VON SCHAEWEN A.
PA Von Schaewen A;
XX WPI: 1999-338905/29.
DR P-PSDB; AAY08890.
DR
XX Nucleic acid encoding plant N-acetylglucosaminyl transferase I
PT useful for generating plants producing glycosylated proteins of low
PT antigenicity
XX
PS Claim 5; Page 27; 37pp; German.
XX
CC This invention describes novel plant N-acetylglucosaminyl transferase I
CC (GnTI) proteins and their encoding nucleic acids. The nucleic acid
CC sequences of the invention may be used for recombinant production of the
CC encoded proteins, which are then used to raise antibodies (Ab) for
CC identifying plants with deficient or defective production of the protein.
CC They may also be used to detect such plants by hybridization and to
CC isolate related sequences from other plants or to generate antisense or
CC sense constructs for reducing/deleting GnTI protein activity in plants.
CC These transgenic plants may be used to produce glycoproteins with
CC minimal, uniform and defined sugar residues. Such glycoproteins are
CC useful in medicine and research, e.g. human glucocerebrosidase for
CC treating Gaucher's disease. Plants which are defective or deficient in
CC production of the GnTI protein can be made to produce glycoproteins with
CC minimal, uniform and defined sugar residues, of low antigenicity. Use of
CC these plants eliminates the need for the difficult isolation and
CC deglycosylation of native proteins or preparation in defective animal
CC cells. This sequence represents cDNA which encodes an Arabidopsis
CC thaliana GnTI homologue.
XX
SQ Sequence 510 BP; 143 A; 104 C; 114 G; 149 T; 0 other;

Query Match 19.7%; Score 329.2; DB 20; Length 510;
Best Local Similarity 77.8%; Pred. No. 9.7e-88;
Matches 397; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 350 atcggagatgtgcagatgccagtgccagctgtagtgtgttattggtcgtcagctgactgac 409
D 1 atcggaaagctggatcccccagtgccgctgtagtgtgttattggtcgtcagctgctgacag 60
QY 410 taacttgagagagactattaaatccattcttaaaataccaaacatctgtttgcatcaaaatat 469
D 61 tatcttgaaaggactgttaaatcagttttaacatacaaaactccgtgtgttccaaatat 120
QY 470 cctcttttcatatcccgatggatcacaatcctctgatgaagaaagcttgccttgagctat 529
D 121 cctctatttatctcaggtggtatctgtatcaagcgtctcaagagcaagtcattgagctat 180
QY 530 ggtcagctgacgtatatgcagcacttggattatgaacctgtgtcactactgaaagaccaggg 589
D 181 aatcaattaacatatatgcagcacttggatttttgaccagctgctcactgaaagcctggc 240
QY 590 gaactggttgcatatacaagattgcacgttcattacaagtggcattggtacgtgcttt 649
D 241 gaactgactgcgtactacaagattgcacgttcactcaacagtggcactggaccagttgtt 300
QY 650 cacaagcataaatttagccgtgttatcactacagagatgatgaaattgctgctgat 709
D 301 tacaacacaaatttagtcagatgattactacagagatgatgaaattgctcagac 360
QY 710 tttttgactatttggagctggagctactcttcttgacagagacagtcgattatggct 769
D 361 ttcttgattactttgagctgcagctagctatctcatggatagggataaaaccattatggct 420
QY 770 attttcttcttggaaatgacaatggacaaaggcagttgctccaagatcctgtactctttac 829
D 421 gcttcacatggactgataatggacagaagcagttgtgtcatgacctatgcgtctatac 480
QY 830 cgtcagacttttttccctgggtcttggatgg 859
```


Human and rabbit DNA sequences encoding GNT I enzymes - for converting mannose to hybrid and complex N-glycan(s)

Claim 3; Fig 4; 57pp; English.

Rabbit liver Gnt I was purified and digested with pepsin, then trypsin. The peptide fragments were sequenced and used to design degenerate PCR primers (see AAQ30220-5). cDNA was prepared from total RNA from rabbit liver. PCR was carried out on the cDNA preparation. One of the two PCR products was cloned into the SmaI site of pGEM7z for sequencing and then used as a riboprobe. The riboprobe was used to screen a rabbit liver cDNA library in lambda gt10. The largest insert in a positive clone was 1.6kb. An 80bp riboprobe was prepared from the 5'-terminal of the 1.6kb insert and used to rescreen the library. The largest cDNA insert was cloned into pGEM-7z to obtain pGEM-7z-rcgntI. The full-length rabbit Gnt I coding sequence was eventually obtained from overlapping clones. See also AAQ25595.

Sequence 2485 BP; 549 A; 623 C; 644 G; 669 T; 0 other:

Query Match 10.9%; Score 181.6; DB 13; Length 2485;
Best Local Similarity 54.1%; Pred. No. 2.4e-43;
Matches 395; Conservative 0; Mismatches 329; Indels 6.

1:

QY	602	tactacaagattgcacgctcattacaagtggcgatcagctgtttccacaagcataat	661
Db	599	tactacaagatcgacgcatcaccgctggcgatcggccaaatctccacaatttccac	658
QY	662	tttagcgtgttatcatatcagaagatgatatggaaattgctgctgatttttgaact	721
Db	659	taccagcagctgtggtgtggaagatgattcgcaggctggcaccagagctcttggag	718
QY	722	tttgaagctggagctactctcttgacagagacaagtcgattatggctattttctctt	781
Db	719	ttcaggcccaattaccacctgttgaaagcagaccctccctctgctgtgtgctgcct	778
QY	782	aatgacaattggcaaaaggcagttctcca-----agatccttgatgcttttacogctca	835
Db	779	aatgacaattggcaaaagacagatgttagactcgagtgaagcagagttactctacog	838
QY	836	gactttttctcgtcttgatggatgctttcaaaatcaacttggctcgaaactatctca	895
Db	839	gattcttcttcggttagctgttactgttgctgaaactctggctgaaactggagccc	898
QY	896	aagtgcccaaggcttactgggatgactggctaaagctgaaagaaaatccacagaggtcga	955
Db	899	aagtgcccaaggcttctgggatgactggatgcgcgcgctgagcagcggaaaggagg	958
QY	956	caatttattcgcccaagatttgcagaacgtacaaatttggtagacatgggtctcagttt	1015
Db	959	gcctgtgctgtccagaatctcagaacaatgacatttgcgggaagggtgtgagccat	1018
QY	1016	ggcgagtttttaagcagatctcttgacgcaaatgaagctaaatgatccaggttgatgg	1075
Db	1019	ggcgagttcttgaccagcatctcaagttcaagctgaaaccagcagtttgtacccttc	1078
QY	1076	aagtcacatggacctaaagttaacttttggaggacaactatgtgaacactttggcgact	1135
Db	1079	accagctggacctgtcgtaccttcagcaggagactatgaccgggatttctctgctcgt	1138
QY	1136	gttaaaaggcttaagcccatccacgagctgactgttttgaagacatttaacatagat	1195
Db	1139	gtttatgtgtctcccagttacaggttggaagatggaggccaatgaccggaaggagcta	1198
QY	1196	ggtgatgtgcgtatctcagtcagagaccaactgaccttgaagatatcgtctcagagtt	1255
Db	1199	ggagaggtgcgctacagtacacaggcagggacagcttcaaggcttctgcgaagccctg	1258
QY	1256	ggcatttttgaagaatggaagatgggtgtaccacgggcagcatataaaggagtagt	1315
Db	1259	ggtgtcatggtacctcaaatcaggctgaccacgggctggaaccgggcattgttccac	1318

Qy 1316 ttccggttcc 1325
 ||| |||
Db 1319 ttcttattcc 1328

RESULT 8

AAI69837
ID AAI69837 standard; cDNA; 1335 BP.

AAI69837:

DT 19-DEC-2001 (first entry)

DE Human beta-1,2-N-acetylglucosaminyltransferase I cDNA.

Human; beta-1,2-N-acetylglucosaminyltransferase; GnT; GnTI;
KW
maltose-binding protein; MBP; fusion; sugar chain synthesis
KW

Homo sapiens.

PN JP2001178453-A.

03-JUL-2001.
PD
AA

24-DEC-1999; 99JP-0368127.

AA
PR 24-DEC-1999; 99JP-0368127.

РА (ТОУМ) ТОУОВО КК.

PA (FUJI/) FUJIYAMA K.

DR WPI; 2001-599905/68.

XX
1005, HQ007034

PT for use in sugar chain synthesis, comprises using *Escherichia coli* - preparation of maltose-binding protein-N-acetylglucosaminyltransferase

PS Claim 5; Page 9-11; 14pp; Japanese.

The invention relates to the preparation of a maltose-binding protein (MBP)-N-acetylglucosaminyltransferase (Gnt) fused protein. The method comprises transforming *Escherichia coli* with an expression vector encoding the fusion protein. Gnt antibody is useful in sugar chain synthesis and in a plasmid pMAL-c2/ECORI-XbaI fragment to give a new plasmid pMGNT-1. The present sequence encodes human Gnt.

SQ Sequence 1335 BP; 215 A; 428 C; 441 G; 251 T; 0 other:

Query Match	9.9%	Score 165.6	DB 22	Length 1335
Best Local Similarity	52.7%	Pred. No. 1e-38		
Matches 385	Conservative	0	Mismatches 339	Indels 6
				Gaps 1

Qy 602 tactacaagattgcacgctcattacaagtqqccatttgaatcagcctgtttcacaaagcataat 661

Db 544 tactacaagatcgcgcgccactaccgcctgcgcgcctgcgcgcaggtcttccggcagtttcac 603

Qy 662 tttagccqtgtttatcatactagaagatgatatggaatgtctactgatttttttactat 721

Db

604 ttccccqccqccgtggtgggtggaggatgacctgagagtggccccgaacttcttcgaatac 663

Qy 722 tttagagcctgagcctactcttcttgacagagacaaatcgaattatagctattctctctgg 781

Db

[illegible][illegible]

836
0v

1000 900 800 700 600 500 400 300 200 100 0

RESULT 10
AAF93806
ID AAF93806 standard; cDNA; 2250 BP.
XX
AC AAF93806;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0120.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
XX
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-0183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI; 2001-093989/11.
DR P-PSDB; AAB88379.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
PT gene therapy or as candidate target molecules in drug development -
XX
PS Claim 1; SEQ ID 125; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretion and
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and
CC activity. The antibodies and antagonists may also be used as therapeutic
CC agents to down regulate expression and activity. The antibodies may also
CC be used as diagnostic agents for detecting the presence of the
CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay
CC (ELISA)). Examples of diseases which may be treated include rheumatoid
CC arthritis and diabetes.
XX
SQ Sequence 2250 BP; 408 A; 561 C; 586 G; 495 T; 0 other;

Query Match 8.6%; Score 144.2; DB 22; Length 2250;
Best Local Similarity 51.7%; Pred. No. 3.3e-32;
Matches 357; Conservative 0; Mismatches 328; Indels 6; Gaps 1;
Oy 641 cagctgtttccacagcataatttttagcgtgtttatctactatagaagatgatgaaatt 700
Db 409 caggcttccgcgcagtttcgtctcccgcccgctggtggtggaggatgacctggagggtg 468

Oy 701 gctgctgatttttttgaactattttgaggtggagctactcttcttgacagagacaagtcg 760
Db 469 gcccggaacttcttcagtaacttccagccacctatccgctctgaaggccgacctctcc 528
Oy 761 attatgctatttcttcttggaatgacaaatggacaaagcagcttcgtccaagat----- 814
Db 529 ctgtgtgctgctcgccctgggaatgacaaagcgaagcagatggtgacgcagcagcag 588
Oy 815 cctgagctcttaccgcctcagactttttccctggtcttcttggtggtggtcttccaaatca 874
Db 589 cctgagctctctacgcgacccgacttttccctggctggctggctgctgtgttggccgag 648
Oy 875 acttggctccgaactatctccaaagtgccaaaggtctactgggatgactggctaaagctg 934
Db 649 ctctggctgagctggagcccaagtggccaaagccttcttggacgactggatgctggcgg 708
Oy 935 aaagaaatcacagaggtcgacaaatttattcggcagaagtttgcagaactgacataattt 994
Db 709 ccggagcagcgagggcgccctgcatacgcctcgagatctcaagaacgatgaccttt 768
Oy 995 ggtgagcatggttctagtttgggagcttttttaagcagatcttctgagccaatgaagcta 1054
Db 769 ggcgcgaagggtgtgagccacggcgagttcttggaccgacacctcaagtattatcaagctg 828
Oy 1055 aatgatgtccaggttgattgggaagtcaatggaccttaagttaaccttttgagagacaactat 1114
Db 829 aaccagcagttgtgcaacttccaccagctggacctgtcttaacctgcagcggaggcctat 888
Oy 1115 gtgaacacatttggcagcttggtaaaaggcttaaacccatccacgagctgctgctgtt 1174
Db 889 gaccgagatttctcccgcgctctacggtgctcccccagctgcagtgagaaagtgaag 948
Oy 1175 ttgaagaacatttaacatagatggtgtgctgtatttcacagacagagaccaactagacttt 1234
Db 949 accaatgaccggaaggagttggggaggttcgggtgcagtatactacggcgaggacagcttc 1008
Oy 1235 gaagatactcgtccagagtttggcatttttgaagaatggaaggtggtgtaccacgggca 1294
Db 1009 aaggcttccgcaaggctcgggtgtcatgagaccttaagtccggggttccgagagct 1068
Oy 1295 gcataataaggagtagtagtttccgggttc 1325
Db 1069 ggtaccggggtattgtcaccttccagttcc 1099
RESULT 11
ABL22693
ID ABL22693 standard; DNA; 2367 BP.
XX
AC ABL22693;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19552.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX

Wed Aug 14 11:48:20 2002

us-09-591-466c-1.rng

Page 16

Search completed: August 13, 2002, 20:33:01
Job time: 8309 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 17:48:37 ; Search time 3105.21 Seconds
(without alignments)
11247.674 Million cell updates/sec

Title: US-09-591-466c-1

Perfect score: 1669
Sequence: 1 gaattcgccgcctgaga.....actttggccgcggaattc 1669

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_othe.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

ALIGNMENTS

RESULT 1

A95053

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

1 (bases 1 to 1669)

Von.S.A.

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTTRANSFERASE I (GNTI)

ACTIVITY

Patent: WO 9929879-A 1 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

Location/Qualifiers

1..1669

/organism="Solanum tuberosum"

/strain="DESIREE"

A95053

Sequence 1 from Patent WO9929879.

A95053

A95053.1

GI:6779205

potato.

Solanum tuberosum


```
/db_xref="taxon:4113"
/cell_type="BLATTZELLEN"
/tissue_type="MESOPHYLL"
/clone_lib="(B) LAMBDA ZAP II (ECORI)"
/dev_stage="M: 'SINK' ORGAN"
15..52
gene
53..1393
/gene="CGL"
CDS
53..1393
/gene="CGL"
/standard_name="GNTI"
/EC_number="2.4.1.101"
/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/notes="ERSTE GNTI-SEQUENZ AUS KARTOFFEL (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE 1"
/protein_id="CAB70462.1"
/db_xref="GI:6779206"
/translation="MGNKPCFDLRYLLVVAALAFIYIQMLRFAATQSEYVDRLLAAIE
AENHCTSORLLLDKLSQOQGVVALEEQKHQDQECRLRALVDLESKGIKKLIGD
VOMPVAQVYMACSRDYLERIKSLKYQTSVASKYPLFIQDGSNPVRLKLSYG
QLTYMHLDYEPVHTERPGELVAYKIARHYKWDLPFKHNFSEVILLEDMEIAA
DFDYPEAGATLLDRDKSMAISSWNDGROFVODPDALYSDPFGGLWMLSKSTW
SELSPKWAKYWDMLRLKENHRGQFIRPEVCRTYNFGEHSSGLOFPKQYLEPIKL
NOVODMKSMDSLYSLLLEDNYKHFGLVKKRPFHGADAVLKAFNIDGSDVRIQYRDQL
DPEDTARQGFIEENKDGVPRAAYKGIIVFRFOTRRRFLVSPDSLRLQIGVEDT"
<80..>139
CDS
/gene="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMBRANANKER (AMINO ACID 10-29)"
/notes="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70463.1"
/db_xref="GI:6779207"
/translation="LRYLLVVAALAFIYIQMLRF"
1394..1655
3' UTR
489 a 312 c 387 g 481 t
ORIGIN

Query Match 100.0%; Score 1669; DB 6; Length 1669;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1669; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattcgagccgctgagaaacctcgaaattcaatttcgcatattggcagagatgagagg 60
Db 1 GAATTCGGCGCGCCTGAGAAACCCCTCGAAATTCATTTTCGCATTTGGCAGAGATGAGAGG 60
QY 61 gaacaagtgttcttgattacgtacctctcgctgctgctgctgctgctgcttcaatc 120
Db 61 GAACAAGTGTTCCTTTGATTTACGCTACCTTCCTGCTGGCTGCTCTCGGCTTCATCTA 120
QY 121 catacagatcgccgttttcgcagacacagtcagaaatagtagaccgcttgcgtgcaat 180
Db 121 CATACAGATCGCGCTTTTCGGCACACAGTCAGAAATATGTAGACCCCTTGCTGCTGCAAT 180
QY 181 tgaagcagaaaaatcattgacaaagtcagaccagattgcttattgacaagattgaccagca 240
Db 181 TGAAGCAGAAAATCATGTGTACAAAGTCAGACCAGCATTGCTTATTGACAAAGATTAGCCAGCA 240
QY 241 gcaagaagagtagtagctcttgaaagcaaaatgaagcatcaggaaccagagagtgccggca 300
Db 241 GCAAGGAAGAGTAGTAGCTCTTGAAGACAAATGAAGCATCAGGACCCAGGAGTCCCGGCA 300
QY 301 attaagggtctctgttcagatcttgaagtaagggcataaaaaagtaaatcgagatgt 360
Db 301 ATTAAGGCTCTTGTTCAGGATCTTGAAGTAGGGGCATAAAAAAGTTAATCGGAGATGT 360
```

```
QY 361 gcagatgccagtgccagctgtagttgtatggcttgagcttcagtcgtagctactacacgtggagag 420
Db 361 GCAGATGCCAGTGCACGCTGTAGTTGTATGGCTTGCAGTCGTACTGACTACCTGGAGAG 420
QY 421 gactattaaatccactctttaaataaccaaaacatctgttgcatcaaaaatctctcttttcat 480
Db 421 GACTATTAAATCCACTCTAAATAACCAACATCTGTTGCATCAAAATATCTCTCTTTTCAT 480
QY 481 atccaggatggatcaaatcctctgatgtaagaagcttgctttagctatagctcagctgac 540
Db 481 ATCCAGGATGGATCAAACTCTGATGTAAAGAAAGCTTGCTTTGAGCTATGCTCAGCTGAC 540
QY 541 gtatatgcagcacttggtattatgaacctgtgcatactgaaagccaggggaactggttgc 600
Db 541 GTATATGCAGCAGCTGGATTATGAACCTGTGCATCTGAAAGACCAGGGGAACCTGGTTGC 600
QY 601 atactcaagattgcacgtcattacaaagtgggcatggatcagctgtttcacaaagcataa 660
Db 601 ATACTCAAGATTGCACGCTATTACAAGTGGCATTTGGATCAGCTGTCTTCCACAAGCATAA 660
QY 661 ttttagccgtgttatcatcactagaagatgatatggaattgctgctgatttttttgaacta 720
Db 661 TTTTAGCCGTGTATCATCTAGAAATATATGGAATTTGCTGCTGATTTTTTTGACTA 720
QY 721 ttttagggctggagctactctctttgacagagacaagtcgattatgctattctctctg 780
Db 721 TTTTAGGCTGGAGCTACTCTCTTGCACAGACAAGTCGATTATGGCTATTTTCTCTTTC 780
QY 781 gaatgacaatggacaaggcagttcgcgaagatcctgatctcttaccgctcagactt 840
Db 781 GAATGACAATGGACAAGGCATTTCTGCAAGATCCTGATGCTCTTTTACCGCTCAGACTT 840
QY 841 ttttctcgctcttgatgagctcttcaaaatcaacttggtccgaactatctccaaagt 900
Db 841 TTTTCTCGCTCTGGATGCTCTTCAAAATCAACTTGGTCCGAACCTATCTCCAAAGTG 900
QY 901 gccaaagcttactgggtagctgagctgaagctgaagaaatacacagaggtcgacaatt 960
Db 901 GCCAAAGGCTTACTGGGATGACTGGCTAAGGCTGAAGAAATAACAGAGGTGCGACAATT 960
QY 961 tattcgcccaagaagtgtgcagaacgtacaattttgtgagcatggttcttagtttgggca 1020
Db 961 TATTCGCCCAGAAAGTTTCAGAACGTCATAATTTGGTGAAGCATGGTCTTAGTTGGGGCA 1020
QY 1021 gtttttaagcagtatctttgagccaattaaagctaaatgatgtccagggtgattggaagtc 1080
Db 1021 GTTTTAAAGCAGTATCTTGAGCCAAATTAAGCTAAATCATGTCTCCAGGTTGATTGGAAGTC 1080
QY 1081 aatgacctaagtcttacccttttggaggacaacatagtgaacaactttggcgacttgggttaa 1140
Db 1081 AATGGACCTAAGTTTACCCTTTTGGAGGACAACATATGTGAAGACACTTTGGCGACTTTGGTTAA 1140
QY 1141 aaaggctaagccccatccacggagctgtagctgttttgaagcatttaacatagatggtga 1200
Db 1141 AAAGCTAAGCCCATCCAGGAGCTGATGCTGTTTTGAAAGCATTTAACATAGATGGTGA 1200
QY 1201 tgtcggtattcagtagacagagacaaactagactttgaagatatcgctcgacagtttggcat 1260
Db 1201 TGTGCGGTATTTCAGTACAGAGACCAACTAGACTTTTGAAGATATCGCTCGACAGTTTGGCAT 1260
QY 1261 ttttgaagaaatggaagatggtgtaccacggcgagcatataaaggagatagttttccg 1320
Db 1261 TTTTGAAGAAATGGAAGATGGTGTACCACCGGGCAGCATATAAAGGGATAGTAGTTTCCG 1320
QY 1321 gtttcaaacatctagacgtgttctctgtttcccttgattctcttcgacaacttgaggt 1380
Db 1321 GTTTTCAAAACATCTAGAGCTGTGTTCTGTTTCCCTTGATTCCTTCTCGACAACCTTGGAGT 1380
QY 1381 tgaagatacttagcagaatgatatgttgagccctgagcaacaatttagacttatttggtag 1440
Db 1381 TGAAGATACTTAGCGAAGATATGATTGGAGCCTTGAGCAACAATTTTAGACTTATTGGTAG 1440
QY 1441 gatacatttgaagaagctgacacgaaaagtatgactaccagtagctacatgcacaacatttt 1500
```

Db	1441	CATACATTTGAAGAGCTGACAGAAAGTATGACTACCACTAGCTACATGCAACATTTT	1500	
Qy	1501	aatgttaatgaaggaaacccactgcttatgttggaatgaatcatcacacacatct	1560	
Db	1501	AATGTTAATGAAGGAACCACTGCTTATTTGTTGAATGGATGAATCATCACACACATCT	1560	
Qy	1561	attattcaagtttacaacataaagaggaaatgttgcctcctataaaaacaaattttttgtt	1620	
Db	1561	ATTATTCAAGTTTACAAACATAAAGAGGAATGTTGCCCTATATAAAACAAATTTTGT	1620	
Qy	1621	tctaagaaggaacgttacgattatgagcaactttggcgccgcgaattc	1669	
Db	1621	TCTAAGAAGCAAGCTTACGATTATGAGCAACTTTGGCGCGCGGAATTC	1669	
RESULT	2			
STU249878				
LOCUS		Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI		
DEFINITION		gene), clone A1.		
ACCESSION		AJ249878		
VERSION		AJ249878.1 GI:18076139		
KEYWORDS		GntI gene; N-acetylglucosaminyltransferase I.		
SOURCE		potato.		
ORGANISM		Solanum tuberosum		
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
TITLE		Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
JOURNAL		1 (sites)		
REFERENCE		Wendroth, I., Tjaden, J. and von Schaewen, A.		
AUTHORS		Isolation and characterization of different plant		
TITLE		N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional		
JOURNAL		analyses in the Arabidopsis cgl mutant, and in potato and tobacco		
FEATURES		antisense plants		
SOURCE		Unpublished		
REFERENCE		2 (bases 1 to 1641)		
AUTHORS		von Schaewen, A.		
TITLE		Direct Submission		
JOURNAL		Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,		
FEATURES		Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,		
SOURCE		GERMANY		
FEATURES		Location/Qualifiers		
gene		1..1641		
CDS		/organism="Solanum tuberosum"		
		/variety="Desiree"		
		/db_xref="taxon:4113"		
		/clone="A1"		
		/tissue_type="leaf mesophyll"		
		/note="sink"		
		39..1379		
		/gene="GntI"		
		39..1379		
		/gene="GntI"		
		/EC_number="2.4.1.101"		
		/function="initiates complex-type glycans in the Golgi"		
		/codon_start=1		
		/evidence=experimental		
		/product="N-acetylglucosaminyltransferase I"		
		/protein_id="CAC80697.1"		
		/db_xref="GI:18076140"		
		/translation="MRGNKFCFLRLVLLVVAALAFIYIQRLFATQSEYVRLAAIAE		
		AEHNKTSOTRLLDKTSQOGRVVALEEMKHODQECRLRALVQDLSEKGIKKLIGD		
		VOMPVAAYVMACSRDYLERTIKSILKYOTSVASKYPLFISODGSPDVRKLLALSYG		
		QLTYMQHLDYEPVHTERPCELVAAYKTAHYKALDQLEFHKHNSRVILLEDMEIAA		
		DFDFYFAGATLLDRKSIKAISSWNDNGORQVQDDALYRSDFFPGLGWLMSKSTW		
		SELSKPWKAYWDWLRLKHNHRGROFIIRPEVCTIRNFHGSSLGOFKQYLEPIKL		
		NDVQWKSMDLSYLLLEDNYVKHFGDLVKKAKPIHGADAVLKAFNIDGDVRYQYRQDL		
		DFEDIAROFGIFEWKDGVPRAYKGIWVRFOTSRVFLVSPDSLRLQIGVED"		
BASE COUNT		485 a 302 c 377 g 477 t		
ORIGIN				

		Query Match	98.3%; Score 1641; DB 8; Length 1641;	
		Best Local Similarity	100.0%; Pred. No. 0;	
		Matches 1641; Conservative	0; Mismatches	0; Gaps
Qy	15	ctgagaaacccctcgaattcgaatttcgcatcttgccagagatgagagggagcaagttttgtc	74	
Db	1	CTGAGAAACCCCTCGAATTCATTTCCCATTTGGCAGAGATCAGAGGGAACAAGTTTCT	60	
Qy	75	ttgatttcagggtaccttctcgtcggtgctctcgccttcacatcacagatgcggc	134	
Db	61	TTGATTTACGGTACCTTCTCGTGGTGGTCTCGCCCTCATCTACATACAGATGCGGC	120	
Qy	135	ttttcgacacagtcagaataatgtagccgcttgctgcgaattgaagcagaaaaac	194	
Db	121	TTTTCCGCACACAGTCAGAAATATGTAGACCGCTTGCCTGCTGCAATTAAGCAGAAAATC	180	
Qy	195	attgtacaagtcagaccagattgcttattgcaaatattagccagcagcaagaagtag	254	
Db	181	ATTGTACAAGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGAGAGTAG	240	
Qy	255	tagctcttgaaagacaaatgaagcatcaggaccaggagtgccggcaattaaagggctctg	314	
Db	241	TAGCTCTTGAAGAACAATAATGAAGCATCAGGACCGAGTGGCGGCAATTAAGGGCTCTTG	300	
Qy	315	ttcaggatcttgaagtaaggcgcataaaaaaattaatcgagatgtgcagatgcagtg	374	
Db	301	TTCAGGATCTTGAAGTAAGGCGCATAAAAAGTTAATCGGAGATGTGCAGATGCCAGTGG	360	
Qy	375	cagctgtagttgttagcttcgcttcagtcgactgactgactgactgactgactgactgact	434	
Db	361	CAGCTGTAGTTGTTATGGCTTGCAGTCGTACTGACTACCTGGAGAGGACTATTAATATCCA	420	
Qy	435	tcttaaaataccaaacatctgttcgacataaaatactcttttcatatcccgagatggat	494	
Db	421	TCTTAAATACCAACATCTGTTGTCATCAAAATATCTCTTTTCATATCCCAGGATGGAT	480	
Qy	495	caaatcctgatgtgaagaagcttgcctttgagctatggtcagctgacgtatagcagcact	554	
Db	481	CAATCTGATGTGAAGAAAGCTTGTCTTTCAGCTATGGTCAGCTGACCTATATGACGACT	540	
Qy	555	tgattatgaacctgtgcatactgaaagaccaggggaaactggtgcatactacaagatg	614	
Db	541	TGGATTATGAACCTGTGCATCTGAAAGACCAGGGGAACCTGGTTGCTACTACTACAAGATTG	600	
Qy	615	cacgtcattacaagtgggcattggatcagctgtttcacagcataattttagcgcgtgta	674	
Db	601	CAGCTCATTTACAAGTGGGCATTGGATCAGCTGTTTCACAAGCATAATTTTAGCCGTGTTA	660	
Qy	675	tcatactagaagatgatattggaattgctgctgattttttgactattttgaggtggag	734	
Db	661	TCATACTAGAAGATGATATGGAATTTGCTGCTGATTTTTCGACTATTTTTCAGGCTGGAG	720	
Qy	735	ctactctttgcagagacaagtcgattatggctattctcttggaaatgacaatggag	794	
Db	721	CTACTCTTTTGACAGAGACAAGTCGATTATGGCTATATTTCTTCCTGGAATGACAATGGAC	780	
Qy	795	aaaggcagttcgtccagaatcctctgctctttcacgcgcagacttttttcctggtcttg	854	
Db	781	AAAGGCAGTTTCGTCGAAGATCTCTGCTCTTTACCGCTCAGACTTTTTCCTGGTCTTG	840	
Qy	855	gatggatgctttcaaatcaacttgctcggaactatctccaaagtggccaaaggttact	914	
Db	841	GATGGATGCTTTCAAAATCAACTTGGTCCGCACTATCTCCAAAGTGGCCAAAGGCTTACT	900	
Qy	915	ggatgaactggtcaagctgaaagaaaaacacagaggtcgacaatttattcgccccagaag	974	
Db	901	GGATGACTGGCTTAGCGTGAAAGAAAAATCACAGAGGTGCACAAATTTATTCGCCAGAGAG	960	
Qy	975	tttcgagaacgtacaatttttgggtgagcaggttctagtttggggcagtttttttaacagt	1034	
Db	961	TTTTCAGAACGTACAATTTTGGTGAGCTGGTTCTAGTTGGGCGCAGTTTTTTTAAACAGT	1020	
Qy	1035	atcttgagccaattaaagctaaatgatgtccaggttgatgggaagtcaatggacctaagtt	1094	

Db	603	ATTGCTGCTGATTTTTTTTTCACATATTTTTCAGGCTGAGCTACTCTTCTTGACAGAGATAAG	662
Qy	758	tcgattatggctattcttcttgggaatgacaatgacaaaggcaggttcgloccaagatccct	817
Db	663	TCGATTATGCTATTTTCTCTTGGGAATGACATGACAAAGCGCAGTTGCTGCAAGATCCT	722
Qy	818	gatgctcttacccgtcagactttttctctgctcttgatgctgtcttcaaaatcaact	877
Db	723	TATGCTCTTTACCGCFCAGACACTTTTTTCTGCTGTGATGATGCTTTCAAAATCAACT	782
Qy	878	tggtccgaactatctccaaagtgcgaagccttactgggatgactggcgaagctgaaa	937
Db	783	TGCTCCGAACATATCTCAAAGTGGCAAGGCTTACTGGGATGACTGGCTAAGGCTCAA	842
Qy	938	gaaatcacagagtgacaaatttatctgccccagagtttcgagaacgtacaaattttgtt	997
Db	843	GAAATCACAGAGGTGCACAAATTTATTCGCCCAGAAAGTTTCAGAACATACAAATTTTGGT	902
Qy	998	gagcatggttctagttg9ggcagttttttaagcagttatcttgagccaattaagctaaat	1057
Db	903	GAGCATGGTCTAGTTTGGGGCAGTTTTTTTAAGCAGTATCTTTGAGCCAATTAAGCTAAAT	962
Qy	1058	gattccaggttatggaagtcaatggacctaaagttaacctttttggggacaactatgtg	1117
Db	963	GATCTCCAGGTTGATTGGAAGTCAATGGACCTTAAGTTACTCTTTTGAGGACAACTATGTG	1022
Qy	1118	aaacactttggcacttgggttaaaaggctaagcccatccacggagctgctgttttg	1177
Db	1023	AAACACTTTGGCGACTTGGTTAAAAAGGCTAAGCCCATCCAGCGTCTGATGCTGTTTG	1082
Qy	1178	aaagcattaaacatagatggtgatgtgcgtattctatcagtcacagacaaactgaactttgaa	1237
Db	1083	AAAGCATTTTAACATPAGATGTGATGTGCGGTATTACGTACAGAGACCACTAGACTTTGAA	1142
Qy	1238	gatactctgcacagttgttggcatttttgaagaatggaaggtggtgtaccacgggcagca	1297
Db	1143	CATATCGCTCGCGAGTTTGGCATTTTGAAGAAATGGAAGGATGGTGTACCACGGGCAGCA	1202
Qy	1298	tataaaggatagtagtttccgggttccaacatctagacgtgtgtctctgtttccct	1357
Db	1203	TATAAAGGGATAGTAGTTTTTCCGGTTTCAACATCTTAGACGTGTTCCTGTGTGCCCT	1262
Qy	1358	gattctctgcacaaacttgagttgaagatacttagcgaagatatga-ttgagcctgag	1416
Db	1263	GATTTCTCTTCACAACTTGGAGTTGAAGATACCTAGCGAAGATATGATTTGGTGCCTGAG	1322
Qy	1417	caacaatttagactatttggtaggatacatatttgaagagctgacacgaaagtatgact	1476
Db	1323	CAACACTTTAGACTTATTTGGTAGGATACATTTGAAGGAGCTGACACGAAAGCATGACT	1382
Qy	1477	accagtagctacatgcacattttaatgtaaggaaggaaaccctactgttattgttga	1536
Db	1383	ACCACTAGCTACATGCACATTTTAATGTTAATGGAAGGAACCTCACTGCTACTGTTGA	1442
Qy	1537	atggatgaatc-atcacacatctctattatcaag-tttacaacataaagaagaatgt	1594
Db	1443	ATGGATGAATCAATCACCACACAGCCTATTATTCAAGTTTTTACAACATAAAGAGG----	1496
Qy	1595	tgccctataaaacaaaattttttgttcttaagaaggaaacttacgattatgagcaacttt	1654
Db	1497	-----GACACATGTTTTTTTCTTAAGAAGGAGCGTTACGATTATGAGCGACTTT	1548
Qy	1655	g 1655	
Db	1549	G 1549	
RESULT	4		
STU249880			
LOCUS	STU249880	1499 bp	mRNA linear
DEFINITION	Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI		
ACCESSION	AJ249880		

VERSION	AJ249880.1	GI:18076143	
KEYWORDS	GntI gene; N-acetylglucosaminyltransferase I.		
SOURCE	potato.		
ORGANISM	Solanum tuberosum		
REFERENCE	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
AUTHORS	1 (sites)		
TITLE	Wenderoth, I., Tjaden, J. and von Schaewen, A.		
JOURNAL	Isolation and characterization of different plant		
REFERENCE	N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants		
AUTHORS	2 (bases 1 to 1499)		
TITLE	von Schaewen, A.		
JOURNAL	Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY		
FEATURES	Location/Qualifiers		
source	1..1499		
	/organism="Solanum tuberosum"		
	/variety="Desiree"		
	/db_xref="taxon:4113"		
	/clone="A8"		
	/tissue_type="tuber"		
	/note="sink"		
gene	37..1158		
	/gene="GntI"		
CDS	37..1158		
	/gene="GntI"		
	/EC number="2.4.1.101"		
	/function="initiates complex-type glycans in the Golgi"		
	/codon_start=1		
	/product="N-acetylglucosaminyltransferase I"		
	/protein_id="CAC80899.1"		
	/db_xref="GI:18076143"		
	/translation="MKRQDECRQLRALVODLESKIKKLIGDVQMPVAVVVMACSR		
	ADYLERIKSLKYQTSVASKYPLFI SQDGNPDVRKRLKLSYDQLTYMHLDFEPVHT		
	ERPELVAYVKIARHYKALDQLFHKHNSRVILLEDMEIAADFDFEAGATLLDR		
	DKSTMATSSNDNGROFQVDPDALYRSDFPGIGWMLSKSTWSELSKPKAYWDDW		
	LRLEKNHRGQFIHPEVCRTYNEGHSLSGQFKOYLRPIKLNDVQVDMKSMDSL		
	LENDYVHKFGDLVKKAKPIHGADAVLKAFNIDGDRIOYRDQLDFAIARQFCIFEW		
	KDGVPRAAKGIIVFRFQTSRRFLVNPDSLRLQGVEDT"		
BASE COUNT	452 a	266 c	339 g 442 t
ORIGIN			
Query Match	81.2%;	Score 1356;	DB 8; Length 1499;
Best Local Similarity	98.5%;	Pred. No. 0;	
Matches 1379;	Conservative	0; Mismatches	20; Indels 1; Gaps 1;
Qy	257	gctcttgaagaacaaatgaagcatcaggaccaggagtgccggcaataaaggctctgtt	316
Db	22	GTTCCTCCGGTAACAAATGAAGCGTCAGGACCGAGTGCCGGCAATTAAGGGCTCTTGT	81
Qy	317	caggatcttgaagtaagggcataaagaattaatcggagatgtgcagatgccagtggca	376
Db	82	CAGGATCTTGAAGTAAGGGCATAAAAAGTTAATCGGAGATGTGCAGATGCCAGTGCA	141
Qy	377	gctgtagttgttattgcttgcagtcgtactgactacacctggagaggactattaaatccatc	436
Db	142	CGTGTAGTTGTTATGGCTTGCGATCGTGTACTTACCTGGAGAGGACTATTAAATCCATC	201
Qy	437	ttaaataaccacaaatctgttgcatacaaaataatcctcttttcataatccagatggatca	496
Db	202	TTAAATAACCAAAACATCTGTTGCATCAAAATATCCTCTTTTCATATCCAGATGGATCA	261
Qy	497	aatcctctgatgaagaagcttgccttgcagctatgcagctgcagctgcagctatgcagacttg	556
Db	262	AATCTGTATGTAAGAAAGCTTCTTTGAGCTATGATGATGACCTGACCTATATGACGACTTG	321

QY	557	gattatgaacctgtgtcactactaaagaccaggggaactggttgctacatactacaagattgca	616
Db	322	GATTTTGAACCTGTGCATACTCAAGACCAGGGAACCTGGTTGCATACTACAAGATTGCA	381
QY	617	cgtcattacaagtgggcattgatcagctggtttccaaagcataaatttttagccgtgttacc	676
Db	382	CGTCATTACAAGTGGGCATTGGATCAGCTGTTTCAACAGCATAAATTTAGCCGTGTTATC	441
QY	677	atactagaagatgatatagaattgctgctgatttttttttttttttttttttttttttttttt	736
Db	442	ATACTAGAAGATGATATGGAANAATGCTGCTGATTTTTTTTGACTATTTTGAGGCTGGAGCT	501
QY	737	actctctttgacagagacaagtcgattatggtctattctcttcttggaaatgacaatggacaa	796
Db	502	ACTCTCTTTGACAGACACAAGTCGATTATGGCTATTCTCTTGTGAATGACAATGGACAA	561
QY	797	aggcaatcgtccaagatcctgaactctcttaccgtcagacttttttccctgctctgga	856
Db	562	AGGCAGTTCGTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTTCTGCTGCTTGA	621
QY	857	tggatgcttttcaaaatcaacttggtcgaaactatctccaaaatggccaaaaggcttactgg	916
Db	622	TGCATGCTTTCAAAATCAACTTGGTCCGAACATATCTCCAAAGTGCCAAAGGCTTACTGG	681
QY	917	gatgactggcctaaggctgaagaataatcacagaggtcgaataattatttcgccaagaagt	976
Db	682	GATGACTGGCTAAGGCTGAAGAAATAACACAGAGGTCGACAATTTATTCACCCAGAAGTT	741
QY	977	tcagaacgtacaaatttggtagcattgctctagtgttggggcagtttttttaagcagtat	1036
Db	742	TGCAGAACGTACAATTTTGGTGAGCATGCTCTAGTTTGGGGCAATTTTTTAAGCAGTAT	801
QY	1037	cttgaagccaaattaagctaaatgatgtccagggttgattggaagtcgaatgacctaagtacc	1096
Db	802	CTTGAGCCCAATTAAAGTAAATGATGTCCAGGTTGATTGGAAGTCATATGACCTAAGTTAC	861
QY	1097	cttttgaagacaaactatgtgaacaacttttggcgaacttgggtttaaagaagcctaagccatc	1156
Db	862	CTTTTGGAGGACAACATATGTGAACAACATTTGGCGACTTGGTTTAAAGAGGCTAAGCCCATC	921
QY	1157	cacggagctgatctgttttgaagcatttaacatagatggttgatgctgcatctcaagtc	1216
Db	922	CACGGAGCTGATGCTGTTTGAAGAGCATTTAAACATAGATGGTGATGGCTGATTTCAGTAC	981
QY	1217	agagacaaactagactttgaagatatcgctcgacagtttggcatttttgaagaatggaag	1276
Db	982	AGAGACCAACTAGACTTTGAAGCTATCGCTCGACAGTTTGGCAATTTTGAAGAAATGGAAG	1041
QY	1277	gatggtgtaccacgpggcagcacaataaagggatagtagttttccgggtttcacaacatctaga	1336
Db	1042	GATGGTGTACCCAGGCGCAGCATATAAAGGGATAGTAGTTTTCGGGTTTCAAAACATCTAGA	1101
QY	1337	cgtgtgtcttcctgttttccctgattctcttcgacaacttggagttgaagatacttagcga	1396
Db	1102	CGTGTGTCTCTGTGTCCTGATTCCTCTCGACAACCTTGGAGTTGAAGATACCTTAGCGA	1161
QY	1397	agatatgattggagcctgagcaacaatttagacttatttggtaggatacatttgaagaag	1456
Db	1162	AGATATGATTGGAGCCTGAGCAACAATTTAGACTTATTTGGTAGGATACATTTGGAAGAG	1221
QY	1457	ctgacacgaaaaatgatgactaccagtagctacatgcaacatttttaattgaagaaga	1516
Db	1222	CTGACACGAAAAGTATGACTTACCAGTAGCTTACATGTCATGCAACATTTTAAATGTTAAT	1281
QY	1517	accactgcttattgttggaaatggaatcatccacacatcctattattcaag-cttac	1575
Db	1282	ACCACCTGCTTATTGTTGGAATGGATGAATCATCATCACCACATCCTATATTCAAGTTTAC	1341
QY	1576	aaacataaagagaaatggttgcctataaaaaacaaatttttttcttaagagggaacgt	1635
Db	1342	AAACATTAAAGAGGAAATGTAGCCCTATAAACACAATGTTTTTGTCTTAAGAAGAACGT	1401
QY	1636	tacgattatgacaaactttg 1655	

NTY16832

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

5'UTR

CDS

3'UTR

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Mismatches

Indels

Gaps

QY

Db

QY

Db

QY

Db

QY

Db

QY

1402

1418

606

427

618

75.4%

Score 1257.8;

DB 8;

Length 2031;

88.8%;

Pred. No. 0;

0;

Mismatches 157;

Indels 20;

Gaps 3;

52

gatgagagggaacaagtttctgttattacggtacaccttcctgctgagctctgc

111

76

GATGAGAGGGTACAAGTTTGTCTGTGATTTCGGGTACCTCCTCATCTTGGCTGCTGTCG

135

112

cttcatctacatacagatcggtcttttcgcgcacacagtcagataatgtagaccgcttgc

171

136

CTTCATCTACATACAGATCGGCTTTTTCGGACACAGTCAGAATATGCAGATCGCCTTC

195

172

tgctgcaattgaagcagaataatcattgtacaagtcagaccagattgcttattgacaagat

231

196

TGCTGCAATTGAAGCAGAAAAATCAGTGTAAAGTCAGACCAGATGCTTTATTGACCCAGAT

255

232

tagccagcagcaaggaaagtagtagtagctcttgaagaacaaatgaagcatcagaccagga

291

|||||
Db 256 TAGCCAGCAGCAAGGAAGATAGTTGCTCTTGAAGACAATAAGAGCTCAGGACCAGGA 315
Qy 292 gtgcggcaataaggacctgttcaggatctcttgaaagtaaggccataaaaaagtttaat 351
Db 316 GTGCCGACAGTTAAGGCGCTCTGTGTCAGGATCTTGAAGATGAAGGCCATAAAAAAGTTGAT 375
Qy 352 cggagatgtcagatgccagtgccagctgtagttggttatggtcttgagtcgactgacta 411
Db 376 CGGAATGTACAGATGCCAGTGCGTGTGTAGTTGTTATGGCTTGCAATCGGCGTACTA 435
Qy 412 cctggaggagactattaaatccatttaaataaccaacatctgttgtcatcaaaatatcc 471
Db 436 CCTGGAAGAAGACTATTAAATCCATCTTAAATACCAATATCTGTTCGCCCAAAATATCC 495
Qy 472 tcttttcataatccaggatgagatacaaatcctgatgtgaagaagcttgcttttgactatgg 531
Db 496 TCCTTTTCATATCCAGATGGATGCATCTGATGTTAGGAAGCTTGCTTTGAGCTATGA 555
Qy 532 tcagctgacgtatcagcagcaacttgattatgaacctgtgtcactacgaaagaccagggga 591
Db 556 TCAGCTCAGCTATATCAGCACCTTGATTTTGAACCTGTGCATCTACTGAAAGACCAGGGGA 615
Qy 592 actggttgcatactacaagattgcagctcattacaagtgggcattggatcagctgtttca 651
Db 616 GCTGATTCATACATAACAAATTCAGCTCATTTACAAGTGGGCATTTGGATCAGCTGTTTA 675
Qy 652 caagcataatttagcgcgtgtatcatactagaagatgatatggaataatgctgctgatttt 711
Db 676 CAACCATAAATTTAGCCGTGTATCATACTAGAGAATGATATGGAATTTGCCCTCATTT 735
Qy 712 ttttgactatttgaggctggagactactctcttgacagagacaagtcgattatggctat 771
Db 736 TTTTGTACTTTTGTAGGCTGGAGCTACTCTCTTTGACAGAGCAAGTCGATATGCGTAT 795
Qy 772 tctcttgaatgaactggacaaaggcagttcgtccaagatcctgatgctctctttaccg 831
Db 796 TCTCTTGTGAATGACAATGGCAATGGCAATGCGATTGTGCCAAGATCCTTATGCTCTTTACCG 855
Qy 832 ctcagactttttccgtgcttgatggatgcttttcaaaatcaactgtgcgaactatc 891
Db 856 CTCTGATTTTTCGGGCTTGGATGGATGCTTTCAAAATCTACTTGGGAGCAACTATC 915
Qy 892 tccaaagtggccaaagcttactggatgactggtcgaagctgaaagaaatccacagagg 951
Db 916 TCCAAAGTGGCCAAAGGCTTACTTGGACACACTGGCTAAGACTCAAAGAGAAATCACAGAGG 975
Qy 952 tcgacaatttattcgcaccagaagtttcgagaacgtacaaatttttgatgacatggttctag 1011
Db 976 TCGACAATTTATTCGCCCAGAAGTTTGCAGATCATATAATTTTGGTGAGCATGGTTCTAG 1035
Qy 1012 tttggggcagtttttaagcagatatttgagccaattaaagctaaatgatgtccaggttga 1071
Db 1036 TTTGGGGCAGTTTTTCAAGCAGTATCTTGAAGCAATTAACATTAATGATGTCCAGGTTGA 1095
Qy 1072 ttggaagtcaatggaccctaaagttaacttttgagagacaactatgtgaacactttggcga 1131
Db 1096 TTGGAAGTCAATGGACCTTAGTTACTCTTTGGAGGACAATTACGTGAAACACACTTTGGTGA 1155
Qy 1132 ctgtgttaaaaggcctaagcccatccacggagctgatgctgttttgaagcatttaacat 1191
Db 1156 CTGTGTTAAAAAGGCTAAGCCCATCCATGGAGCTGATGCTGTTTGAAGAGCATTTAAGAT 1215
Qy 1192 agatgtgtgatgtgcgtattcagtagacagacaactagaacttttgaagatatcgtctgaca 1251
Db 1216 AGATGCTGATGGCTATTACGTACAGAGATCAACTAGACTTTTGAAGATATCGCACGGCA 1275
Qy 1252 gtttggcatttttgaagaatggaaggatggtgtaccacggcgacatataaaaggatagt 1311
Db 1276 ATTTGGCATTTTGAAGAAATGGAAGGATGGTGTACCAGGGGCACCATATAAAGGAATAGT 1335
Qy 1312 agtttccggtttcaaacatcagaagtggttctcctgtttccctctgattctcttcgaca 1371
|||||

Db 1336 GGTTTTCGGGTACCAACGCTCAGACGCTGATTCTCTTGTGGCCCTGATTGCTTCAACA 1395
Qy 1372 acttggagttgaagatacttagcgaagatatgattgagcctgagcaacaatttagactt 1431
Db 1396 ACTCGGAATGAAGATACTTAACAAGATATGATTGGAGCCCGGACAAAGATTTAGACTT 1455
Qy 1432 atttggtagatacatatttgaagagctgcacgaaaaagtatgac-----ta 1477
Db 1456 ATTGGGTAGGATGCATCGAGCTGACACCAACCATGAGTTTACCAGTTACATACAACGTT 1515
Qy 1478 ccagtagctacatgcacaacttttaattgaaggaagaccctcttattgtt-gga 1536
Db 1516 TTAAATTGTTATGGAGGAGCTCAGTCTTCTAGTGTGTAAGGGATATCGGCTCTTAATA 1575
Qy 1537 atggtgaatcaccacacatcctattt-----caagtttacaacataaagagaaa 1591
Db 1576 TTGGATGAATCATCACAACTATTTTAAAGCAAGTGTTCGGAACATAAAGAGAAA 1635
Qy 1592 tgttgccctataaaaaaa 1610
Db 1636 TGTAGCCCTGTAAGACAA 1654

RESULT 6

NTA249883
LOCUS
DEFINITION
1708 bp mRNA linear PLN 04-JAN-2002
Nicotiana tabacum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A9.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
NTA249883.1 GI:18076147
AJ249883
GntI gene; N-acetylglucosaminyltransferase I.
common tobacco.
Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (sites)
Wenderoth, I., Tjaden, J. and von Schaewen, A.
Isolation and characterization of different plant
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional
analyses in the Arabidopsis cgl mutant, and in potato and tobacco
antiseNSE plants
Unpublished
2 (bases 1 to 1708)
von Schaewen, A.
Direct Submission
Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,
GERMANY

FEATURES

Location/Qualifiers
1..1708
/organism="Nicotiana tabacum"
/variety="Samsun NN"
/db_xref="taxon:4097"
/clone="A9"
/tissue_type="leaf mesophyll"
/note="sink"
113..1453
/gene="GntI"
113..1453
/gene="GntI"
/EC_number="2.4.1.101"
/function="initiates complex-type glycans in the Golgi"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAC80702.1"
/db_xref="GI:18076148"
translation="MRGNKFCDFRYLLILAAVAFIYTMRLFATOSEYADRLAAATE
AENHCSQTRLLIDQLSLOQGRIVALEBQMKRODCRQLRALVQDLESGIKKLIGN
VQMPVAOVVMACNRRADYLEKTKILKYQISVASKYPLFISDGGSHPDVRLKALSYD
OLTYMOHLDFPEVHTERPGLIAYKIAIRHYKHALQDLFYKHNFSRVILDDMEIAP
DEPDFEAGATLLDRKSKMAISSWNDNGOMQVDPYALYRDSFPFGLGWLMSKSTW
DELSPAWPKRAYMDDWLRLKENHNRGQFIRPEVCRTYNFGEHSSLSQGFQKYLEPIKL

NDVOVDKSMDSLVSLLDNYVYKHFGDLVKKAKPIHCADAVLKAFFNIDGVRIOYRDQL
DFENIARFGFIFFSEKHKGVPRAAYKGIYVFRYQTSRRVFLVGHDSLQQLGIEDT*
BASE COUNT 506 a 326 c 380 g 496 t
ORIGIN

Query Match 75.2%; Score 1255; DB 8; Length 1708;
Best Local Similarity 89.1%; Pred. No. 0;
Matches 1412; Conservative 0; Mismatches 140; Indels 33; Gaps 4;

QY 52 gatgagagggaacaagtgtttgcttgattacggtacccttctcgtggtgctctcgc 111
|||||
Db 112 GATGAGAGGGAACAAGTGTTCGTGCTGATTCCGGTACCTCCTCATCTTGGCTGCTGCG 171
|||||
QY 112 cttcatctacatacagatgcgcttttccgcacacagtcagaatatgtagaccgccttgc 171
|||||
Db 172 CTTTCATCTACACAGATCGCGCTTTTTCGCACACAGTCAGAAATATCGAGATCGCCTTGC 231
|||||
QY 172 tctgcaattgaagcagaaaaatcattgtacaagtcagaccagattgcttattgacaagt 231
|||||
Db 232 TCGTGAATTGAAGCAGAAAAATCATGTACAGCCAGACAGATTGCTTATTGACCAGAT 291
|||||
QY 232 tagccgcagcagaagaagttagtgcctctgaagacaaaatgaagcatcagagaccagga 291
|||||
Db 292 TAGCCTGCACGACGAAGAAGTATGCTCTTGAAGAACAAAATGAAGCGTCAGGACCAGGA 351
|||||
QY 292 gtccgcgaattaaaggtcttctgttcaggaatcttgaagaagggcataaaaaagtta 351
|||||
Db 352 GTCCCGACAAATTAAGGGCTCTTGTTCAGGATCTTGAAGTAAGGGCATAAAAAGTTGAT 411
|||||
QY 352 cggagatgtcgagatgccagtgccagctgtagtgtgtatggttcagtcgtactgacta 411
|||||
Db 412 CGGAATGTACAGATGCCAGTGGCTGCTGTAGTTGTATGGCTTGAATCGGCTGATTA 471
|||||
QY 412 cctggaggagactataaatccatctttaaataaccaaacatatctgttgcatacaaaatcc 471
|||||
Db 472 CCTGGAAGAGACTATTAATCCATCTTAAATACCAATATACCAATATCTGTCGCTCAAAATATCC 531
|||||
QY 472 tcttttcatacccagatggataaactcctgatgtaagaagcttgccttgagctatgg 531
|||||
Db 532 TCTTTTCATATCCCAGATGGATACATCCTCATGTCAGGAAGCTTGCCTTGAGCTATGA 591
|||||
QY 532 tcagctgacgtatgacgaccttggtattatgaacctgtcatactgaagcagagcaggga 591
|||||
Db 592 TCAGCTGACGTATATGCAGCACTTGGATTTTGAACCTGTGCATACCTGAAGACCAGGGA 651
|||||
QY 592 actggtgcatactacaagatgcacgtcattcacagtggtggcattggatcagctgttca 651
|||||
Db 652 GCTGATTGCATACTACAAAATTCGACGTCAATTACAAGTGGGCATTGGATCAGCTGTTTA 711
|||||
QY 652 caagcataattttagccgtgttatcatatacagatgatatggaattgctgctgattt 711
|||||
Db 712 CAAGCATATTTTACCGTGTATTATCATACTAGAAGATGATATGGAATTTGCCCTTGATTT 771
|||||
QY 712 ttttgactattttgagctggagctactcttctgacagacaaatcagatcattatgctat 771
|||||
Db 772 TTTTGACTTTTTTGGCTGGAGCTACTCTCTTGACAGACACAGTCGATATATGCCTAT 831
|||||
QY 772 tctcttggaaatgcaatggcaagcagtcgtctccaagatcctgatgctctttaccg 831
|||||
Db 832 TTTCTCTTGGAAATGACAATGGACAAATGCAGTTTGTCTCAAGATCCTTATGCTTTTACC 891
|||||
QY 832 ctccagacttttctgctgttggatgcttccaaatcaaatcacttggctccgaactatc 891
|||||
Db 892 CTCAGATTTTTTCCCGTCTTGGATGGATGCTTTCAAAATCTACTTGGGACGAATATTC 951
|||||
QY 892 tccaaatggccaaaggcttactggatgactggctgaagcgtgaaagaaatcacagag 951
|||||
Db 952 TCCAAAGTGGCCAAAGGCTTACTGGGACGACTGGCTTAAGACTCAAGAGAAATCACAGAG 1011
|||||
QY 952 tcgacaattttatcccccagaagtttgcagaacgctacaattttggtgagcattggttctag 1011
|||||
Db 1012 TCGACAATTTATTCCGCCAGAAAGTTTGACAGACATATAATTTTGTGTGAGCATGGTCTTAG 1071
|||||

QY 1012 tttggggcagcttttttaagcagtatctcttgagccaatttaagctaataatgatgtcccaggttga 1071
|||||
Db 1072 TTTGGGCGAGTCTTTTCAAGCAGTATCTTGAGCCAATTAATTAATGATGTCCAGGTTGA 1131
|||||
QY 1072 ttgaaagtcaatgacctaaagtacccttttggaggacaaactatgtgaaacactttgcga 1131
|||||
Db 1132 TTGAAGTCAATGAGCTTGTAGTTACCTTTTGGAGGACAAATACCTGGAACACTTTTGGTGA 1191
|||||
QY 1132 cttgggttaaaaggctaaagccctccacggagctgtagctgtgttttgaagcatttaacat 1191
|||||
Db 1192 CTTGGTTTAAAGAGCTTAAGCCCATCCATGAGCTGATGCTGTCTTGAAGCATTTAAACAT 1251
|||||
QY 1192 agatggtgattgcgtattcagtcagagaccacacagctagacttttgaagatatcgtctcaga 1251
|||||
Db 1252 AGATGCTGATGTGCGTATTTCAGTACAGAGATCAACTAGACTTTTGAAATATTCGCACGGCA 1311
|||||
QY 1252 gtttgcatttttgaagaatggaagatggtgtaccacgggcagcatataaaaggatagt 1311
|||||
Db 1312 ATTTGGCATTTTGAAGAATGGAAGGATGGTGTACCACGTGCACCATATAAAGGAATAGT 1371
|||||
QY 1312 agtttccgggttcaaacatctagacgtgtgttccctgtttccctctgattctcttcgaca 1371
|||||
Db 1372 AGTTTTCGGTACCAAAACGTCAGACGTGATTCTTGTGGCCATGATTGCTTCAACA 1431
|||||
QY 1372 acttgagttgaagatactattagcgaagatatgatt---ggagcctgagcacaacatttga 1428
|||||
Db 1432 ACTCGGAATTTGAAGTACTTTAACAAGATATGATTCAGAGCGCCGGCAAAATTTTGA 1491
|||||
QY 1429 cttatttgttagatcacatttgaagagctgcacagaaagtagtactaccagtagctac 1488
|||||
Db 1492 CTTATTGGGTAGTAGTCATCGAGCTGACACTAACCATGATTTT-----ACCAAGTTAC 1544
|||||
QY 1489 atgcaacattttaatgttaatggaaggaacccactgct-----t 1527
|||||
Db 1545 ATACAACGTTTTAATGTTATACGGAGGAGCTACTGTTCTAGTGTTCGAAGGATATCGGC 1604
|||||
QY 1528 attgttggaaatggaatgaatcaccacacacctattat--ttcaagttttacaacataaag 1585
|||||
Db 1605 TTTCTTAGTATTGGATGAATCATCAACACACCTATTATTATTAAAGTGTTCAGAACATAAG 1664
|||||
QY 1586 aggaatgttgcctcataaaacaa 1610
|||||
Db 1665 AGGAATGTAGCCTGTAAAGACTA 1689
|||||

RESULT 7
A95055
LOCUS A95055 1737 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO9929879.
ACCESSION A95055
VERSION A95055.1 GI:6779208
KEYWORDS common tobacco.
SOURCE
ORGANISM
Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.
Von,S.A.
REFERENCE
AUTHORS 1 (bases 1 to 1737)
TITLE VEGETABLE GntI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
A REDUCED OR LACK OF N-ACETYLGUCCOSAMINYLTRANSFERASE I (GntI)
JOURNAL Patent: WO 9929879-A 3 17-JUN-1999;
VON SCHAEWEN ANTJE (DE)
FEATURES
source
1. 1737
/organism="Nicotiana tabacum"
/strain="SAMSUN NN"
/db_xref="taxon:4097"
/cell_type="BLATTZELLEN"
/tissue_type="MESOPHYLL"
/clone_lib="LAMBDA ZAP II (ECORI)"


```
/clone_lib="(B) CLON(E): GNTI-A9(T)"
/dev_stage="M: 'SINK' ORGAN"
15..126
127..1467
/genes="CGL"
127..1467
/genes="CGL"
/PC_number="2.4.1.101"
/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/notes="ERSTE GNTI-SEQUENZ AUS TABAK (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGALUCOSAMINYLTRANSFERASE I"
/protein_id="CAB70464.1"
/db_xref="GI:6779209"
/translation="MRGNKFCDFRYLLILAAVFIYQMRLFATQSEYADRLAAIE
AENHCTSTRLLIDQISLQOGRIVALEQMRKQDOECRLRALVDLSEKGIKLIGN
VQMPAAVVVWACNRADYLEKTIKSLKQISVASKYPLFIQDQSHDPVKRLALSYD
QUTYMHQDLDFEPVHTERPEGLIAYKIKARHYKWDLDOLFYKHNFESRVIILEDMEIAP
DFFDFEAGATLLDRDKSISMAISSWNDNGOMQVODPYALYRSDFPGLGWLMSKSTW
DELSPKWPAYWDDMLRLKENHGRGOFIRPEVCRTYNEGEHSSLGQFFKQYLEPIKL
NDVOYDMKSMDSLYLLENYKHNFGRGOFIRPEVCRTYNEGEHSSLGQFFKQYLEPIKL
DPENIARQFGIFEEWKGVPRAAYKGIWVFRYQTSRRVFLVGHDSLQQLGIEDT"
<154..>5213
CDS
/genes="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMBRANANKER (AMINO ACID 10-29)"
/notes="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70465.1"
/db_xref="GI:6779210"
/translation="FRVLLILAAVAFIVQMRLF"
1468..1723
3'UTR
BASE COUNT 510 a 337 c 390 g 500 t
ORIGIN
```

```
Query Match 75.2% Score 1255; DB 6; Length 1737;
Best Local Similarity 89.1% Pred. No. 0;
Matches 1412; Conservative 0; Mismatches 140; Indels 33; Gaps 4;

Qy 52 gatgagagggaacaagtttcttgatttaccggttaccttctcgtcgtggtcgtctcgc 111
Db 126 GATGAGAGGGAACAAGTTTCTGCTGATTTCCGGTACCTCCTCATCTCTGGCTGCTGCG 185

Qy 112 ctctcatcacaacagatcggtcttcttcgacacagtcagaatgatagaccgcttgc 171
Db 186 CTTTCATCTACACACAGATGCGGCTTTTTCGACACAGTCAGAAATATGACAGATCGCCTTGC 245

Qy 172 tctgcaattgaagcagaaatcattgacaagtcagaccagattgcttattgacaagat 231
Db 246 TGCTGCAATTGAAGCAGAAAATCATTTACAAAGCAGACACAGATTGCTTATTGACACAGAT 305

Qy 232 tagccagcagcaggaagagtagctcttctgaagacaaatgaagcatcagacacagga 291
Db 306 TAGCCTGCAGCAGAGAGATAGTTCCTTTGAAGAACAAATGAAGCGTCAGGACACAGGA 365

Qy 292 gtgcgcgaattgaaggtcttcttcaggatcttgaagtaagggcataaaaaagtaaat 351
Db 366 GTGCCGCAATTGAAGGCTCTTCTCAGGATCTTGAAGATGAAGGCATATAAAGTTGAT 425

Qy 352 cggagatgcaatcccaatgagcgttagtctgttattgcttgcgtcgtactgacta 411
Db 426 CGGAATGTACAGATCCAGTGGCTGCTGTAGTGTATAGGCTTGAATCGGCTCATTA 485

Qy 412 ccttgagaggactataatccatcttaaaataccaaacatctgttgcatcaaaaatccc 471
Db 486 CCTGGAAAAGACTATTAAATCCATCTTAAATACCAAAATATCTGTTGGCTGCAAAATATCC 545
```

```
Qy 472 tctttcatatccagagatgatcaaatcctgatgtaaagaaactgtcttgagctatgg 531
Db 546 TCTTTTCATATCCAGGATGGATCACATCCTGATGTCAGAAAGCTTGTCTTGAGCTATGA 605

Qy 532 tcagctgacgtatatatcagcacttggattatgaacctgtgcatactactgaagaccagggga 591
Db 606 TCAGCTGACGTATATGACGACTTGGATTTTGAACCTGTGCATCTGAAAGACACAGGGGA 665

Qy 592 actgggttcataactacaagatgcagctcattacaagtgggcatggatcagctgtttca 651
Db 666 GCTGATTGCTACTACAAAATTGCACGTCATTACAAAGTGGGATTTGGATTCAGCTGTGTTTA 725

Qy 652 caagcataatttagccgtgttatcatactagaagatgatgtgaaatgtcgtcgtatit 711
Db 726 CAAGCATAAATTTAGCCGTGTTATCATACTAGAAGATGATATGGAATTTGCCCTGATTT 785

Qy 712 ttttgaactattttaggctggagctactcttctgaagagacaagtgcattatggctat 771
Db 786 TTTTGACTTTTTTGAGGCTGGAGCTACTCTTCTGACAGAGACAAGTCGATTTATGGCTAT 845

Qy 772 ttctcttgggaatgacaaatggacaaaggcagttcgtccaagatcctgatcgtctttaccg 831
Db 846 TTCTTCTTGGGAATGACAATGGACAATCGAGTTTGTCCAAGATCCTTATGCTCTTTTACCG 905

Qy 832 ctcaagctttttctcgtgcttggatgagatgctttcaaaatcaactggtccgaactatc 891
Db 906 CTCAGATTTTTTCCCGGCTTGGATGGATGCTTTCAAAATCTACTTGGGACGAATTTATC 965

Qy 892 tccaaagtggccaaaggcttactgggatgactggcctaaggctgaagaaataacacagagg 951
Db 966 TCCAAAGTGGCCAAAGGCTTACTGGGACGACTGGCTTAAGACTCAAGAGAAATCACAGAGG 1025

Qy 952 tcgacaattttatccgccagaagtttgcagacgtacaaattttggtagcatggtctctag 1011
Db 1026 TCGACAATTTATCCGCCAGAAAGTTTGCAGACATATAAATTTGCTGAGCATGCTTCTAG 1085

Qy 1012 ttggggcagttttttaaagcagatcttgaagcacaatgaagtaaatgatgtccaggttga 1071
Db 1086 TTTGGGCGAGTTTTTCAAGCAGATATCTTGAGCCAATTAACATAAATGATGTCCAGGTTGA 1145

Qy 1072 ttgaaagtcgaatggacctaaagttaaccttttggaggaacactatgtgaacactttgcga 1131
Db 1146 TTGAAAGTCAATGGACCTTAGTTACCTTTTGGAGAGACAATTTACCTGAAACACTTTTGTGA 1205

Qy 1132 ctgtgttaaaaggctaagccatccacgagcgtgatcgtgttttgaagcatttaacat 1191
Db 1206 CTTGGTTAAAGAGGCTAAGCCCATCCATGGAGCTGATGCTCTTGAAGACATTTTAACAT 1265

Qy 1192 agatgggtgatgtgcgtatttcagtcacagagaccacactagactttgaagatatcgtcga 1251
Db 1266 AGATGGTGATGTGCGTATTTCAGTACAGAGATCAACTAGACTTTTGAAATATCGCAGCGCA 1325

Qy 1252 gtttggcattttgaagaatggaaggtggtgtaccacgggcagcatataaaggatagt 1311
Db 1326 ATTTGGCATTTTGAAGAATGGAAGGATGGTGATCCACGTCGACGATATAAAGGAATAGT 1385

Qy 1312 agtttccggttttcaaacatctagacgtgttctcgttttccctgattctcttcgaca 1371
Db 1386 AGTTTTCCGGTACCAAAAGCTCCAGCGTGATTTCTTGTGGCCATGATTTCCGCTTCAACA 1445

Qy 1372 acttggagttgaagatactctagcgaagatatgatt---ggagcctgagcaacaatttga 1428
Db 1446 ACTCGGAATTGAAGATACTTAACAAAGATATGATTGACAGAGCCCGGCAAAATTTTGA 1505

Qy 1429 ctatttctagatatacatttgaagagctgcacagaaagatgactaccagtagctac 1488
Db 1506 CTTATTGGGTAGATGCATCGAGCTGACACTAAACCATGATTTT-----ACCAGTTAC 1558

Qy 1489 atgcaacattttaatgttaattggaaggaacccactgct-----t 1527
Db 1559 ATACAACGTTTTTAATGTTATACGGAGGAGCTCAGCTGTTTCTAGTGTGAAGGGATATCGGC 1618

Qy 1528 attgttggaaatggaatcatcaccacatcctatta---ttcaagtttacaacaataaag 1585
```


Db	1619	TTCTTAGTATTGGATGAATCATCAACACACCTATTATTTTAAGTGTTCAGAACATAAAG	1678
Qy	1586	aggaaatgttcctcataaaaaacaa	1610
Db	1679	AGGAAATGTAGCCCTGTAAGACTA	1703
RESULT	8		
NTA249882			
LOCUS	Nicotiana tabacum mRNA for N-acetylglucosaminoltransferase I (GntI gene), clone A4.	1836 bp	linear
DEFINITION	Nicotiana tabacum mRNA for N-acetylglucosaminoltransferase I (GntI gene), clone A4.		PLN 04-JAN-2002
ACCESSION	AJ249882		
VERSION	AJ249882.1	GI:18076145	
KEYWORDS	GntI gene; N-acetylglucosaminoltransferase I.		
SOURCE	common tobacco.		
ORGANISM	Nicotiana tabacum		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridaceae; euasterids I; Solanales; Solanaceae; Nicotiana.		
AUTHORS	1 (sites)		
TITLE	Wenderoth, I., Tjaden, J. and von Schaewen, A.		
JOURNAL	Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1836)		
TITLE	von Schaewen, A.		
JOURNAL	Direct Submission		
FEATURES	Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY		
source	Location/Qualifiers		
	1..1836		
	/organism="Nicotiana tabacum"		
	/variety="Samsun NN"		
	/db_xref="taxon:4097"		
	/clone="A4"		
	/tissue_type="leaf mesophyll"		
	/note="sink"		
gene	27..1367		
	/gene="GntI"		
CDS	27..1367		
	/gene="GntI"		
	/EC_number="2.4.1.101"		
	/function="initiates complex-type glycans in the Golgi"		
	/codon_start=1		
	/product="N-acetylglucosaminoltransferase I"		
	/protein_id="CAC80701.1"		
	/db_xref="GI:18076146"		
	/translation="MRGYPCDDFRYLLILAAVAFTYIQMLRFATSEYADRLAAAEI AENHCTSRFLTDIQLSQOGRIVALEQMKRQDQECRLALVQDLESKGTKKLIGN VQMPAAVVMACNPRADYLEKTIKSLKYOISVAPKPLFIQDQSHPDVRKIALSYD QLTVMOHLDIPEVHTPEGLIAYYKIARHYKWLDDOLFVKHFSRVIILEDMEIAT DFRDFEAGATLLDRDKSMALSSNDGOMQFQODPYALYRSDFPGLGWLMSKSTW DELSPKPKAYKAWDLRLKENHRGOFIRPEVCRSYNFGHGSSUGGFQFKYLEPIKL NDVQVDMKSHDLSYLEDNAYRHFGDLVKKAFIHGADAVLKAFNIDGVRIOYRDQL DFDSDIARQGIIEEWMQDGPVRAAYGVFVRYQTSRRFLVGPDSLQQLGNEDT"		
BASE COUNT	536 a	354 c	403 g
ORIGIN			
	Query Match	74.7%;	Score 1246.8; DB 8; Length 1836;
	Best Local Similarity	91.1%;	Pred. No. 0;
	Matches 1359; Conservative	0;	Mismatches 127; Indels 6; Gaps 3;
QY	52	gatgagaggaacaaagtttgcgtttgatattacggtacccttcgctgctgctcctcgc	111
Db	26	GATGAGAGGTTACAAAGTTTTCGTGTGATTTCCCGTACCTCCTCATCTTGCTGCTGCG	85
QY	112	cttcatctacatacagatcgcgcttttcgcacacagtcagaatatctagaccgccttcg	171

Db	86	 CTTCATCTACATACAGATGGGGCTTTTGGACACACAGTCAGAATATGCAGATGCGCTTGC	1415
Qy	172	tgtctgaattgaagcagaaataatcttacaagtccagaccagatgtcttatttgcacaagt	231
Db	146	TGCTGCAATTGAACGACAGAAATCACTGTACAACTGACACGACAGATTGCTTATTGCACGAT	205
Qy	232	tagccagcagcaaggaagtagtagctcttgaagacaataatgaagcatcagagaccagga	291
Db	206	TAGCAGCAGCAAGGAAGAATAGTTGCTCTTGAAGAACAAATCAAGCGCTCAGGACCAAGA	265
Qy	292	qtccgcgaactaaaggctctgttcagagctcttgaagtaaggcataaaaaagtttaac	351
Db	266	GTGCCGACAGTTAAGGCTCTGTGTACAGATCTTGAAGATTGAAGGATGAAGGATAAAAAGTTGAT	325
Qy	352	cggagatgtcagatgccagtgccagtcgttagttgttatggcttgcagtcgtactgacta	411
Db	326	CGGAAATGTCAGATGCCAGTGCGTGTAGTTGTTATGGCTTGCAATCGGGCTGACTA	385
Qy	412	octggagaggaactataataaccatcttaaaataccaaacatctgttgcatacaaaataacc	471
Db	386	CCTGGAAAGAACTATTAAATCCATCTTAAATACCAAAATATCTGTTCGCCAAAAATATCC	445
Qy	472	tctttcatccccagatggatcaaatccctgatgaaagaaagcttgccttgagctatgg	531
Db	446	TCTTTTTCATATCCACGATGGATCACAATCCTGTATGTTAGGAAGCTTGCTTTGAGCTATGA	505
Qy	532	tcagctgacgtatatgcagcaacttgattatgaacctgtgcatactgaaagaccaggga	591
Db	506	TCAGCTCAGCTATATGCAGCACCTTGGATTTTGAACCTGTGCATCTGAAAGACCAGGGCA	565
Qy	592	actggtgcatactacaagattcacgctcaatacaagtgggcatggatcgaactgtttca	651
Db	566	GCTGATTGCACTACAAAAATGCACGTCAATTACAAGTGGCATTTGGATTCAGCTGTTTGA	625
Qy	652	caagcataatttagccgctgtatcatactagaagatgatatggaaattgctgctgattt	711
Db	626	CAAGCATAAATTTAGCCGTATTATCATACTAGAAGATGATATGGAATTTGCCCTTGATTT	685
Qy	712	ttttgactatttgaggctgaactactcttctgcagagacaagtcgattatagctat	771
Db	686	TTTTGACTTTTTGAGGCTGGAGCTACTCTCTTTGACAGACAACTGCGATTATGGCTAT	745
Qy	772	ttctcttggaaatgcaatggacaaaggcagttcgtccaaagatccctgatgctcttaccg	831
Db	746	TTCTTCTTGAATGCAATGGACAAATGCAGTTTGTCCAAGATCCTTATGTCTCTTTACCG	805
Qy	832	ctcagactttttccctggctcttgatggatgctttcaaaatacaactggccgaactatc	891
Db	806	CTCTGATTTTTTTCCCGCGCTTGGATGGATGCTTCCAAATCTACTTGGGACGAACATC	865
Qy	892	tccaaagtggccaaaggcttactgggatgactggctgaagcgtgaagaaaaatccagagg	951
Db	866	TCCAAAGTGCCAAAGGCTTACTGGGACGACTGGCTAAGACTCAAAGAAATCACAGAGG	925
Qy	952	tcgacaatttatcgccaggaagttgcagacaagctacaaatttggctgagcatggtcttag	1011
Db	926	TGCACAAATTTATTGGCCCAAGATTTGCAGATCATATAATTTTGTGTGAGCATGTTCTTAG	985
Qy	1012	tttggggcagtttttaagcagtatcttgagccaattaaagctaaatgatgtccaggttga	1071
Db	986	TTTGGGGCAGTTTTTCAACGAGTATCTTGAGGCCAATTAACATAATATGATGTCCAGGTTGA	1045
Qy	1072	ttggaagtcaatggacctaaagttaacctttggaggacaactatgtgaacaactttggcga	1131
Db	1046	TTGGAAAGTCAATGGACCTTAGTTACCTTTTGAGGACACAATACGTTGAAACACTTTTGTGTA	1105
Qy	1132	cttgggttaaaaggctaaagcccatccagagactgatctattttgaagcatttaacat	1191
Db	1106	CTTGGTTTAAAAGGCTTAAGCCCATCATGGAGCTGATGCTGTTTGTGAAGCATTTTACAT	1165
Qy	1192	agatggtgatgtcgtatttcagtaagagaccaactagacttttgaagatatctgcctgaca	1251

Db	1166	AGATGGTGATGTCGATTTCACAGAGATCAACTAGACTTTGTAAGATATATCCGACGGCA	1225
Qy	1252	gtttggcattttgaagaatgaagatggtgtaccacggcgagcatataaaggatagt	1311
Db	1226	ATTTCGGCATTTTGAAGATGCGAGATGGTGTACACGGGCGAGCATATAAAGGAATAGT	1285
Qy	1312	agtttccgggttccaaacatctagacggtgtgtctctgtttccccctgattctcttcgaca	1371
Db	1286	GGTTTTCGGTACCAAAAGCTGCAGACGTGTATTCTCTGTTGGCCCTGATTTCGCTTCAACA	1345
Qy	1372	acttgaggttgaagatacttagcgaagatatgattggagcctgagcaacaatttagactt	1431
Db	1346	ACTCGGAATGAAGATACTTACAAGAATATGATGTTGGAGCCCGGACAAAGATTTAGACTT	1405
Qy	1432	atttggtaggatacatttgaagagctgacacgaaagttagctaccagtagctacatg	1491
Db	1406	ATTGGTAGGGCA---TGCATCGAGCTCACACCAACCATGAGTT--TACCAAGTTACATA	1460
Qy	1492	caacatttttaa-tgttaatgaagaacccccactgcttattgttggaaatggat	1542
Db	1461	CAACGTTTTTAATTGTATATGAGGAGCTCACTGTTCTAGTGTCTGAAGGGAT	1512
RESULT	9		
ATH243198			
LOCUS		1820 bp mRNA linear PLN 24-AUG-1999	
DEFINITION		Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I (GlcNAcT-I gene).	
ACCESSION		AJ243198	
VERSION		AJ243198.1 GI:5139334	
KEYWORDS		GlcNAcT-I gene; glycosyl transferase; N-acetylglucosaminyltransferase I.	
SOURCE		thale cress.	
ORGANISM		Arabidopsis thaliana	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
AUTHORS		1 (bases 1 to 1820)	
TITLE		Bakker,H., Lommen,A., Jordi,W., Stiekema,W. and Bosch,D.	
JOURNAL		An arabidopsis thaliana cDNA complements the	
MEDLINE		N-acetylglucosaminyltransferase I deficiency of CHO lec1 cells	
REFERENCE		Biochem. Biophys. Res. Commun. 261 (3), 829-832 (1999)	
AUTHORS		99373163	
TITLE		2 (bases 1 to 1820)	
JOURNAL		Bakker,H.	
COMMENT		Direct Submission	
FEATURES		Submitted (17-JUN-1999) Bakker H., Molecular Biology, Centre for Plant Breeding and Reproduction Research, P.O. Box 16 Wageningen, 6700AA, NETHERLANDS	
source		Related sequence AL035538.	
gene		Location/Qualifiers	
CDS		1..1820	
		/organism="Arabidopsis thaliana"	
		/variety="Colombia"	
		/db_xref="taxon:3702"	
		/clone="PABE-AtGN1"	
		/clone_lib="PABE-A.t.seliq"	
		31..1365	
		/gene="GlcNAcT-I"	
		31..1365	
		/gene="GlcNAcT-I"	
		/EC_number="2.4.1.101"	
		/function="glycosyltransferase"	
		/codon_start=1	
		/product="N-acetylglucosaminyltransferase I"	
		/protein_id="CB45521.1"	
		/db_xref="GI:5139335"	
		/translation="MARISCDURFLIPAAMFIYIQMRLFOTQSOYADRLSSAIESE	
		NHCVSMRGLDEVSTQSRIVALEDMKNRQDEELVOLKDLIQTFKKGTAKLTQCGQ	
		MFTAAVVMNACSADYLERTKVSVLTQTPVASKYPLFIQSDGSDQAVKSKLSYNOL	
		TYMQHLDPEPVYTERPGLTAYKTAHYKQWALDQLFYKHKSRVILIEDDEIAPDF	
		EYDFFAASLMDRDKTMAASSNDNQKQFVHDPIALYSDFPFPGWMLKRSTWDE	
		LSPKPKATWDWLRLKENHGSKQFIRPEVCTRYNFGHGSSLGQSFQSLPIKILND	

	VTVDWKARDKLYLTGEGNTTKYFSLGVROARPIQGSDLVLKAQNIKDDVRIRYRKDQVEE	
BASE COUNT	532 a 342 c 421 g 525 t	
ORIGIN	ERIAEGFIPEEMKDGVPRATYKGVVVFRITQTTRRVFLVGPDSDVMQLGRNS*	
Query Match	45.1%	Score 752.6; DB 8; Length 1820;
Best Local Similarity	72.7%	Pred. No. 5.2e-179;
Matches 971; Conservative	0; Mismatches 364; Indels	0; Gaps 0;
QY	65 aaagtttgccttgattacggtacaccttcgctcgctggctcctcgccctcatcacata 124	
Dd	37 AGCATCTCGTGTGACTTCAGATTCTTCTCATCCGGCAGCTTTCATGTTTCATCTACATC 96	
QY	125 cagatgcggcttttcgcagacacagtcagaatatatgaaccgccttgtcgtgcgaattgaa 184	
Dd	97 CAGATGAGGCCTTTCCAGACGCAATCACAGATATGCATGCGCTTCAGTTCGGCTATCGAA 156	
QY	185 gcagaaaatcatgttacaagtgcagaccagatgtcttatgacaagatagccagcagcaa 244	
Dd	157 TCTGAGAACCATTGCACCTAGTCAATCGAGGCCCTCATAGATGAAGTTAGCATCAAACAG 216	
QY	245 gaaagtagtagctcttgaaagaacaaatgaagcatcaggaccaggaggtccggcaatta 304	
Dd	217 TCGCGGATTTGTGCCCTCGAAGATATGAAGAACGCCAGGACGAAGAACTTGTCAGCTT 276	
QY	305 agggcctctgttcaggatcttgaagtaaggccaataaaaaagttaaatcgagatgtgcag 364	
Dd	277 AAGGATCTTAATCCAGACGTTTTGAAAAAAAAGGAATAGCAAACTCATCAAGGTGGACAG 336	
QY	365 atgccagtggcagctgtagttgttatggcttgcagtcgactgaactacctggagaggaact 424	
Dd	337 ATGCTGTGGCTGCTGTAGTGTGTTATGGCTGTCAGTCTGCGACACTATCTTGAAGGACT 396	
QY	425 attaaatccatcttaaataaccaaacatctgtgtcatcaaaatatcctctttcatatcc 484	
Dd	397 GTTAAATCAGTTTTAAACATATACAACCTCCGCTGCTTCAAAATATCCTCTATTATATCT 456	
QY	485 caggatgatcaaatcctgatgaagaagctgtcttggactatggtcagctgaagctat 544	
Dd	457 CAGGATGATCTGATCAAGCTGTCAAGAGCANAGTCATTGAGCTATATAATCAATTAACATAT 516	
QY	545 atgcagcacttggattatgaacctgtgcatactgaaagaccaggggaactggttgcatac 604	
Dd	517 ATGCAGCACTTGGATTTTGAACCAAGTGGTCAC TGAAAGGCTTGGCGAAGCTGACTGCGTAC 576	
QY	605 tacaagattgcacgtcattcacagtgggcattggatcagctgtttcacaaagcataattt 664	
Dd	577 TACAAGATTGCAGCTCACTACAAGTGGGCACTGGACAGTTGTTTACAAACACAATTT 636	
QY	665 agccgtgtatcactacagaagatgatgaaatgctgctgatcttttgaactattt 724	
Dd	637 AGTCGATGATTACTAGAAAGATGATGSGAAATTGCTCCAGACTTCTTTGATTACTTT 696	
QY	725 gaggtggagctactctttgcacagacaaatcgattatggctatttcttcttggaaat 784	
Dd	697 GAGGCTCAGCTAGTCTCATGGATAGGGATAAAACCAATTATGCTGCTTCATCATGGAAT 756	
QY	785 gacaatggacaaagcagttctccaagatccctgatgctcttttacogctcagactttttt 844	
Dd	757 GATAAATGGACAAGACAGTTTGTGCATGATGCCATATGCCGTATACCAGTACAGATTTTTT 816	
QY	845 cctggcttggatggatgctttcaaatcaacttggccgaactatctccaaglgggcca 904	
Dd	817 CCTGGCCTTGGTGGATGCTCAAGAGATCGACTTGGGATGAGTTATCAACCAAGTGGCCA 876	
QY	905 aaggcttactggatgactggcttaaggctgaaagaaaaatcacagaggtgcgcaatttt 964	
Dd	877 AAGGCTTACTGGGATGATTGGCTGAGACTAAAGGAAAAACCAATAAAGGCGCGCAATTCAT 936	
QY	965 cgcccagaatttgcagaacgtacaattttgttgagcatggttcttagtttggggcagttt 1024	
Dd	937 CGACCCGAAGCTGTGAGACATACAATTTGGTGAACATGGGCTCTAGTTTGGGACAGTTT 996	

QY	1025	ttaagcagatctttagccaattaagctaaatgatgtccaggttgattgaagcaatg	1084
Db	997	TTCACTCAGTACTCTGGAACCTATAAAGCTAAACGATGTACCGTGTACCTGGAACGAAG	1056
QY	1085	gaactaagttaaccttttagagagacaactatgtgaacacctttggcagcttggttaaaaag	1144
Db	1057	GACCTGGGATACCTGACAGAGGGAAACTATACCAAGTACTTTCTGGCTTAGTGAGACAA	1116
QY	1145	gctaagcccatccacgagctgatgctgttttgaagcatttaacatagatgggtgatg	1204
Db	1117	GCACGACCAANTCAAGGTTCCTGACCTTGCTTTAAGGCTCAAAACATATAAGGATGATGTT	1176
QY	1205	cgtattcgtacagagaccaactagactttgaagatatcgctcgacagtttggcatttt	1264
Db	1177	CGTATCCCGGTATAAAGACCAAGTAGAGTTTGAACGCATTGACGGGAATTGGTATATTT	1236
QY	1265	gaagaatgaaagatggtgtaccacgagcagcatataaaggatagtagtttcccgatt	1324
Db	1237	GAAGANTGGAAGGTGTGTGCCCTCGAACACAGCATATAAAGGAGTAGTGGTGTTCGAATC	1296
QY	1325	caaacatcagacgtgtgttctctgtttcccttgattctcttcgacaacttgagattgaa	1384
Db	1297	CAGACACACAGACGTGATTTCTGTTGGGCCAGATTCTGTAAATGCAGCTTGAATTCGA	1356
QY	1385	gatacttagcgaaga	1399
Db	1357	AATTCCTGATGCAAA	1371
RESULT	10		
LOCUS	ATH249881		
DEFINITION	Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I (GntI gene).	1830 bp	mRNA linear
ACCESSION	AJ249881		PLN 04-JAN-2002
VERSION	AJ249881.1		
KEYWORDS	GntI gene; N-acetylglucosaminyltransferase I.		
SOURCE	thale cress.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	Wenderoth, I., Tjaden, J. and von Schaewen, A.		
TITLE	Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1830)		
AUTHORS	von Schaewen, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY		
FEATURES	Location/Qualifiers		
source	1. .1830		
	/organism="Arabidopsis thaliana"		
	/variety="Columbia"		
	/db_xref="taxon:3702"		
gene	/tissue_type="whole plant"		
	117..1451		
	/gene="GntI"		
CDS	117..1451		
	/gene="GntI"		
	/EC_number="2.4.1.101"		
	/function="initiates complex-type glycans in the Golgi"		
	/codon_start=1		
	/product="N-acetylglucosaminyltransferase I"		
	/protein_id="CAC80700.1"		
	/db_xref="GI:18072833"		
	/translation="MARISCDLRLFLIPAAFWFIYIQMLRFQTSQYADRLSSAIESE		

BASE COUNT	540 a	348 c	422 g	520 t
ORIGIN				
Query Match	44.8%	Score 747.8;	DB 8;	Length 1830;
Best Local Similarity	72.5%;	Pred. No. 8.4e-178;		
Matches	968;	Conservative	0;	Mismatches 367;
			Indels	0;
			Gaps	0;
QY	65	aagtttgcttggattacggtacgttctcgtcgtgctgctcgtcgttccttcacatcata	124	
Db	123	AGGATCGTGTGACTTTCCTCTCATCCCGCAGCTTTCATGTTTCATCTACATC	182	
QY	125	cagatcggtgttttcgcgcacacagtcagaatatgtagaccgcttgcgtgctgcaattgaa	184	
Db	183	CAGATGAGGCTTTTCCAGACGCAATCACAGTATGCAGTCGCCTCAGTTCGCTATCGAA	242	
QY	185	gcagaaatcattgtacaagtcagaccagattgcttattgacaagattagccagcgcaa	244	
Db	243	TCGAGAACCATTCGACTAGTCAATCGCAGGCTCATAGATGAAGTTAGCATCAACAG	302	
QY	245	ggaagagtagtgcctctgaagaacaaatgaagcatcaggaccagagagtgccgcgaattca	304	
Db	303	TCGCGGATTTGTCCTCGAAGATATGAAGACCGCCAGCAGCAGCAAGCTTGTGCAGCTT	362	
QY	305	agggctctgttcagatcttgaagtaaggcagcataaaaaaagttaatcgagatgacag	364	
Db	363	AAGGATCTAATCAGACGTTTGAAAAAAGGAATAGCAAAACTCTACTCAAGGTGCACAG	422	
QY	365	atgccagtgccgctgtagttgttgccttcagtcgctcgtactgactcactcctcgagagact	424	
Db	423	ATGCTGTGGCTGCTGTAGTGCTTATGGCTCAGCTCGTGCAGACTATCTTGAAGGACT	482	
QY	425	attaaatcaatcttaaaataccaaacatctgttgcatcaaaaatactcttttcatactc	484	
Db	483	GTTAAATCAGTTTAAACATATCAAACTCCCGTGTGTTCAAAATATCTCTATTATCT	542	
QY	485	caggatgatcaaatcctgatgaagaagcttgccttgagcctatggtcagctgacgtat	544	
Db	543	CAGGATGATCTGATCAAGCTGTCAGAGCAAGTCAATTGAGCTATATCAATTAACATAT	602	
QY	545	atgcagcacttggattatgaacctgtgcatactgaaagaccaggggaactggttgcatac	604	
Db	603	ATGCAGCACTTGGATTTTGAACCACTGCTCACTGAAAGCCCTGGTGAAGTGCCTGCTAC	662	
QY	605	lacaagattgcacgtcattacaagtggcattggatcagctgttttcaacaacaataattt	664	
Db	663	TACAAGATTGCAGCTCACTACAAGTGGCACTGGACACTGGACAGTTGTTTACAAACACA	722	
QY	665	agccgtgtatcatcactagaagatgatgaaattgctgcgtgatttttttgactatttt	724	
Db	723	AGTCAGTGATTAATAGAGAGCATATGCAAAATTGCTCCAGACTTCTTTGATTACTTT	782	
QY	725	gagcctgagctactcttttgacagagacaagtcgattatgctatttttcttcttgaat	784	
Db	783	GAGCTGCAGCTAGTCTCATGGATAGGGATAAAACCATTATGGCTGCTTCATCATGAAT	842	
QY	785	gacaatggacaaaggcagttcgtccaagatcctgactgctctttaccgctcagactttttt	844	
Db	843	GATAATGGACAGAAGCAGTTTGTGATGATCCCTATGCGGTATACCGATCAGATTTT	902	
QY	845	ccggctcttgatggatgctttcaaatcaacttggctcgaactatctccaaagtggcca	904	
Db	903	CTGGCCCTGGGGTATGCTCAAGAGATGACATTGGGATGAGTTATCACCAAAAGTGCCA	962	
QY	905	aagccttactgggatgactggctaaaggctgaagaataacacagagagtcgacaattatt	964	

```
Db 963 AAGCTTTACTGGGATGATTGGCTGAGACTAAAGAGAAACCATAAAGGCCGCCCAATTCATT 1022
Qy 965 cgccagaaagtttgcagaaacgtacaaatttttggtgagcatggttcttaagtttggggcaagttt 1024
Db 1023 GCACCGGAAGTCTGTAGAACATACAATTTTGGTCAACATGGGTCTAGTTTGGGACAGTTT 1082
Qy 1025 tttaacgagtattctgaagcaataaagctaaatgatgtccaggttgattggaagtaagt 1084
Db 1083 TTCACTCAGTATCTGGNACCTATATAAGCTAAACGATGTGACGGTTGACTGGAAGCAAG 1142
Qy 1085 gacctaaagttacatttttggaggacaactatgtgaaacacttttggcgacttgggttaaaaaag 1144
Db 1143 GACCTGGGATACCTGCAGAGAGGAAACATATACCAAGTACTTTTCTTGCTTAGTGAGACAA 1202
Qy 1145 gctaagcccatccagagcgtgatgctgttttgaagcattttaacatagatgtagtg 1204
Db 1203 GCAGACCAATTCAGGTTCTGACCTTGTCTTAAGGCTCAAAACATAAAGATGATGAT 1262
Qy 1205 cgtattcagtcagagaccactagactttgaagatatacgtcgacagtttggcattttt 1264
Db 1263 CGTATCCGTTATAAGACCAAGTAGAGTTTGAACGCATTGCAGGGGAATTTGGTATATTT 1322
Qy 1265 gaagaatggaagtagtggttacccagggcagcatataaaagggatagtagttttccggttt 1324
Db 1323 GAAGAATGGAAGGATGCTGTGCCACGACAGCATATAAAGAGAGTAGTGGTGTTCGAATC 1382
Qy 1325 caaacctagacgtgttccctgttccctgattctcttcgacaacttgaggttga 1384
Db 1383 CAGACACAGACGCTGATTCTCGTGGGCCAGGATCTGTAAATGCAGCTTGGNAATCGA 1442

Qy 1385 gatacttagcgaaga 1399
Db 1443 AATTCCTGATGCAAA 1457

RESULT 11
LOCUS A95057
DEFINITION Sequence 5 from Patent WO929879.
ACCESSION A95057
VERSION A95057.1 GI:6779211
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Von S. A.
TITLE VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GntI)
ACTIVITY
JOURNAL Patent: WO 929879-A 5 17-JUN-1999;
VON SCHAEWEN ANTJE (DE)
FEATURES
    source
        1..1854
            /organism="Arabidopsis thaliana"
            /strain="COLOMBIA"
            /db_xref="taxon:3702"
            /tissue_type="ALLE GEWEBE"
            /clone_lib="LAMBDA UNIT-ZAP (ECORI/XHOI) UND LAMBDA ACT"
            /clone_lib="(XHOI)"
            /clone_lib="(B) CLON(E): PBSK-ARA-GNTI-FULL #8"
            /dev_stage="M: REIFE PFLANZEN"
            19..134
            135..1469
            /gene="CGL"
            135..1469
            /gene="CGL"
            /standard_name="GNTI"
            /EC_number="2.4.1.101"
            /function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
            GLYKOPROTEINEN"
```

```
/note="ERSTE GNTI-SEQUENZ AUS ARABIDOPSIS (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"
/protein_id="CAB70466.1"
/db_xref="GI:6779212"
/translation="MAR1SCDLRFLLIIPAFMFYIQMRLFQTSQYADRLSIAIESE
NPTSOMRGLIDEVSIKQSRIVALDEDMKNRQDEELVOLKDIOTFFKKGIAKLQTGGQ
MPVAAVVVWACSRADYLERIVKSVLTQYTPVASKYPLFIQSDGSDAOKVSKLSYNOL
TYWQHLDFEPVYTERGELTAYKIKARHYKWLADOLFYKHKESRVLILEDMETAPDF
FDFAEASLMDRDKTIIMASSNNDNGQKQFVHDPTALIRSDRFFGLGWLKLRSTWDE
LSPKPKAYMDWLRLKENHKQFIAPVCRTPIFGEHSSLGQFFSQFLRYPKLMD
VYDWKAKDLGYLTGNTYKFSGLVRQARPIQGSDLVKLAQNIKDDRIYRKQVEF
ERLAGEFGIFEKWKDGPRTAYKGVVVVFRIQTRRVFLVGPDSVMQLGIRNS"
<157..>215
/gene="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMRANKER (AMINO ACID 8-27)"
/note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=2
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70467.1"
/db_xref="GI:6779213"
/translation="EISSHPGSHVHLHPDEAF"
1470..1848
3'UTR 545 a 357 c 430 g 522 t
BASE COUNT
ORIGIN

Query Match 44.8%; Score 747.8; DB 6; Length 1854;
Best Local Similarity 72.5%; Pred. No. 8.4e-178;
Matches 968; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

Qy 65 aagtttgcttaattacggtaccttctcgctgctgctctcgctctcatcatcata 124
Db 141 AGGATCTCGTGAAGTTCTTCTCATCCCGCAGCTTTCATGTTTCATCTACATC 200
Qy 125 cagatgcggttttccgcagacagtcagaataatgacccgcttctgctgcaattgaa 184
Db 201 CAGATGAGGCTTTTCCAGACGCAATCACAGTATGCAGATCGCCTCAGTTCGCTATCGAA 260
Qy 185 gcagaaaaatcattgtacaagtcagaccagattgcttattgacaagattagccaagcaaa 244
Db 261 TCTGAGAACCATTCGACTAGTCAAAATCGAGCGCTCATAGATGAAGTTAGATCAAAACAG 320
Qy 245 ggaagagtagtagctcttgaagaacaaatgaagcatcagaccagaggtgcggcaatta 304
Db 321 TCGCGGATTGTTGCCCTCGAAGATATGAGAACCGCCAGGACGGAAGAACTTGTGCAGCTT 380
Qy 305 agggctcttctcaggatcttgaaagttaagggaataaaagttaatacggagatgtagcag 364
Db 381 AAGGATCTAATCCAGACGCTTTGAAAAAAGGAATAAGCAAAACTCAGTCAAGGTGGACAG 440
Qy 365 atccagtcgagctgtagtttattgcttcgagtcgactgactactacctggagagagact 424
Db 441 ATGCCTGTGCTGCTGTAGTGGTTATGCGCTGCGAGTCTGCGAGACTATCTTGAAGGACT 500
Qy 425 attaatccatctttaaataccaaacatctgttgcatcaaaataatctcttttcatatcc 484
Db 501 GTTAAATCAGTTTTAACATATCAAACTCCCGTTGCTTCAAAATATCTCTATTATATCT 560
Qy 485 caggaatgatacaaatcctctgatgtaagaaagcttgccttgagctatgtagctgacgtat 544
Db 561 CAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTCATTGAGCTATAATCAATTAACATAT 620
Qy 545 atgcagcacttgattatgaacctgtgtcatcaciagaagaccaggggaactggttgctaac 604
Db 621 ATGACGACTTGGATTTTGAACAGTGGTCTCCTCAAAAGGCTTGGTGAAGTCTGACTGCGTAC 680
Qy 605 tacaagaattgcagctcattacaagtggtgcatgttgatcagctgttttcacagcataatttt 664
Db 681 TACAAGATTGCAGTCACTACAAGTGGGCACATGACCAGCTTGTTTTCACAAACACAAATTT 740
```

QY	665	agcgtgttatcatactagaagatgatatggaattgctgctgatttttttgactatttt	724
Db	741	AGTCGAGTGATTATACTAGAAGACGATATGGAATTTGCTCCAGACTTCTTTGATTACTTT	800
QY	725	gaggctggaagctactcttctgaacagacaagctgatatgctattctctcttggaaat	784
Db	801	GAGCTGCAGCTAGTCTCATGAGTAGGGATAAAACCATTAATGCGCTATACCGATCAGATTTTTTT	860
QY	785	gacaatggacaaaggcagttcgtccaaagatcctgtgctctttaccgctcagacttttt	844
Db	861	GATAATGGACAGAAGCAGTTTGTGCATGATCCCTATACCGCTATACCGATCAGATTTTTTT	920
QY	845	cttggtcttgatgtagtcttccaatacaacttgctgcgaactatctccaagtggcca	904
Db	921	CCTGGCCTTGGGTGGATGCTCAAGAGATCGACTTGGGATGAGTTATCACCAAAAGTGCCA	980
QY	905	aagccttactggatgactgctgaagcgtgaaagaaaaatacacagaggtcgacaatttatt	964
Db	981	AAGCCTTACTGGGATGATTGGCTGAGACTAAAGGAAACCATAAAGCGCCCAATTTCATT	1040
QY	965	cgccagaagtcttgacagaacgtacaattttggtgagcatggttctagtttggggcagttt	1024
Db	1041	GCACCGGAAGTCTGTAGAACATACAATTTTGGTGAACATGGGTCTAGTTTGGGACAGTTT	1100
QY	1025	ttaagcagctatcttgagccaattaagctaaatgatgtccaggttgatttggaaagtcgaatg	1084
Db	1101	TTCACTCAGTATCTGGNACCTATAAAGCTAAACGATGTGACGGTTGACTGGAAGCAAG	1160
QY	1085	gacctaaagttaccttttgagggaacaactatgtgaaacaactttggcgacttgggttaaaaag	1144
Db	1161	GACCTGGGATACCTGCAGAGGAAACTATPACCAAGTACTTTTCTGGCTTAGTCAGACAA	1220
QY	1145	gctaagcccatcacagagctgctgttttgaagcatttaacatagatgctgctg	1204
Db	1221	GCAGCACAATTCAAGTCTGACCTTGCTTTAAAGCTCAAAACATAAAGGATGATGAT	1280
QY	1205	cgtattcagtcacagagaccactagactttggaagatatcgctgcagactttggcattttt	1264
Db	1281	CGTATCCGGTATAAAGACCAAGTAGAGTTTGAACGGCTTCAGGGGAATTTGGTATATTT	1340
QY	1265	gaagaatggaaggatggtgtaccagggcagcatataaaaggatagtagtttccggttt	1324
Db	1341	GAAGAATGGAAGGATGTTGTGCCACGACACATATAAAGAGGATAGTGGTTCGGAATC	1400
QY	1325	caaacatctagacgtgttcttctgtttccctgtattctctcttcgcacaacttggagttgaa	1384
Db	1401	CAGACAACAGACGTGTATTCTCTGGTTGGCCACAGATTCTCTAATGCAGCTTGGAAATCGA	1460
QY	1385	gatacttagcgaaga	1399
Db	1461	AATTCTCTGATGCAAA	1475
RESULT 12			
MUSTRANSX			
LOCUS	MUS musculus (clone E31.1 in pGEM7zf(+))		
DEFINITION	N-acetylglucosaminyltransferase I mRNA, complete cds.		
ACCESSION	L07037		
VERSION	L07037.1 GI:202145		
KEYWORDS	N-acetylglucosaminyltransferase I.		
SOURCE	Mus musculus (strain BALB/c, sub_species domesticus) (library: cDNA of Robert Larsen) cDNA to mRNA.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	Kumar,R., Yang,J., Eddy,R.L., Jr., Byers,M.G., Shows,T.B., Jr. and Stanley,P.M.		
TITLE	Cloning and expression of the murine gene and chromosomal location of the human gene encoding N-acetylglucosaminyltransferase I		
JOURNAL	Glycobiology 2, 383-393 (1992)		

MEDLINE		93043841	Location/Qualifiers	
FEATURES			source	
			1. .2695	
			/organism="Mus musculus"	
			/strain="BALB/c"	
			/sub_species="domesticus"	
			/db_xref="taxon:10090"	
			/cell_line="F9 embryonal carcinoma"	
			/tissue_lib="stem cell"	
			/tissue_lib="cDNA of Robert Larsen"	
CDS			364. .1707	
			/codon_start=1	
			/product="N-acetylglucosaminyltransferase I"	
			/protein_id="AAA40478.1"	
			/db_xref="GI:202146"	
			/translation="MLKQTAGLVLWGALIFVGNALLLLFFWTRPAPGRRLSDSALGDDPSALTREVIHLAEADAEALERQLQOIKEHYALMQRWRVPTVAPAMPVRPVT	
			PSPVQIPIIVACDRSTRRCLDLHYRPSAERFP IIVSDCQHEETAPVAVTASYGTA	
			VTHIQPDLNIAVOPDHRKFOGYKIHRYRWALQIENKFKFPAAVVEDLEVP	
			DFEVEFOATYPLLRTPDPSLMCVSAWNDGKEQWDSKPELLYRTDFPGLGLMLLAD	
			LNAELEKHPKAFWDDMWERPEQRKGRACIRPEISRTMTFGRKGVSHGQFDOHLKFI	
			KLNQOFVPTQLDLSYQEAAYDRFLAQYVGAPOLOVEKYRTNDOKELGEVRVOYTS	
			RDSFKAFKALGVMDLKSVPFRAGTIGTIFQFRRRVHLAPPQTWTGYDPSWN"	
terminator			1705	
3'UTR			1708. .2695	
polyA_site			2695	
BASE COUNT		546 a	715 c	780 g 654 t
ORIGIN				
Query Match		12.1%	Score 202.4;	DB 10; Length 2695;
Best Local Similarity		55.9%;	Pred. No. 2.6e-40;	
Matches		408;	Conservative 0;	Mismatches 316; Indels 6; Gaps 1;
QY	602	tactacaagattcacagctcaattacaagtgaggcattgagatcagctgtttccaagaataat	661	
Db	913	TACTACAAGATTGCCAGGCACCTACCGCTGGCAGCTAGGCCAGATCTTCAACAAGTTCAAG	972	
QY	662	tttagccgtgttatcatactactagaagatgatatggaattgctgtgattttttgactat	721	
Db	973	TTCCCGGCCCTGTGGTGTGGAGGATGATCTGGAAGTGGCCACGACTTCTTTTGAGTAC	1032	
QY	722	tttgagcgtgagcactcttcttgacagagacaaagtcgattatgctattcttcttgg	781	
Db	1033	TTCCAGCGCCACCTACCCACCTGCTGAGAACACAGACCCCTCCCTTTGGTGTGTCTGTCTGG	1092	
QY	782	aatgacaatggacaaaggcagttcgtcca-----agatcctgatgctctttaccgctca	835	
Db	1093	AATGACAAATGGCAAGGAGAGATGGTAGACTCAAGCAAACTGAGCTGCTCTATCGAACA	1152	
QY	836	gactttttcctggttcttgatggtgcttttcaaaatcaaaacttggtccgaactatctcca	895	
Db	1153	GACTTTTTTCTGGCCTTGGATGCTGCTTGTGGTGTATCTGTGGCAGAGCTGGAGCCC	1212	
QY	896	aagtggccaaaggcttactggatgactggtgaagcgtgaaagaaatacacagagtcga	955	
Db	1213	AAGTGGCCCAAGGCTTTTGGGACGACTGGATGCGCCGACCTGAGCAGCGGAAGGACGG	1272	
QY	956	caattatttcgcccaagaagtttgacagacgtacaaattttggtgagcagtggttctagtgtt	1015	
Db	1273	GCTTGATTCTGCAGAAATTTCAAGAACTATGACCTTTGGTCGCAAGGCTGTGAGCCAC	1332	
QY	1016	gggcagtttttaagcagatctcttgaccgaataataagcataatgatgtccagggttgattg	1075	
Db	1333	GGGCAGTCTCTTTGACCAAGCATCTTAAGTTTCATCAAGCTGAACACGACGAGTTCGTC	1392	
QY	1076	aagtcaatgacctaagttaaccttttgaggagacaactatgtgaacaactttggcgacttg	1135	
Db	1393	ACCCAGTTGGACCTGCTGCTACCTGCAGCAGAGGCCCTATGACCCGGGACTTCTTCGCC	1452	
QY	1136	gttaaaaggctgaagcccatccacggagctgctgtgttttgaagcatttaacatagat	1195	
Db	1453	GTCTATGGTCCCCCCCAGCTACAGGTGGAGAAAGTAAGGACCAATATCATCAAGAAGAG	1512	

Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu.
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A..

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 17 Row: p Column: 22
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 202145.

FEATURES

Location/Qualifiers
1..2651
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:11538 IMAGE:3969155"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI.CGAP_Mam1"
/lab_host="PH10B"
/note="Vector: PCMV-SPORT6"
300..1643
/codon_start=1
/product="Similar to mannoside
acetylglucosaminyltransferase 1"
/protein_id="AAH06629.1"
/db_xref="GI:14919415"
/translation="MLKQTAGLVWGAIIFVGNWALLLLFFWTRPAPGRLPDSALG
DPASLTREVIHLADEAEERLGLQIQKEHYALQRWRVPTAPPAPRPVPT
PSPQIPLVIACDSTVRCLDKLLHSPSAERPIIIVSDCQHEETATYASGTA
VFTHIQPDLNTAVQPDHRKFGQYKIAHRYWALGQIFNFKFPAAVVVEDLLEVP
DFEYFQATYPLTRDPSLWCSAWNDKQEMVSDSKPELLYRTDFPGLGLWLLAD
LWAELEKWPKAFMDWRRPEQRKACIRPEISRTWTFCFKGSHGQFEDQHLKFI
KLNQFVPTQLDLSYLOQEAENDFLAQYGAPOLOVEKVTNDQKELGEVRYVYTS
RUSFKAFKALGVMDLKSQVPRAGYRGIVTFQFRGRVHLAPPTWTGYDPSWN"

CDS

RESULT 15
AL606829
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AL606829 267628 bp DNA linear HTG 17-JAN-2002
Mus musculus chromosome 11 clone RP23-10M12, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.
AL606829 AL606829.12 GI:18250804
HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Rodentia; Sclurognathi; Muridae; Mus.
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
Chapman, J.
Direct Submission
Submitted (16-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk
On Jan 19, 2002 this sequence version replaced gi:17384519.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bM10M12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 262703 bases at least Q40
Consensus quality: 264321 bases at least Q30
Consensus quality: 265684 bases at least Q20
Insert size: 266728; sum-of-contigs
Insert size: 174833; 23.7% error; agarose-fp
Quality coverage: 10.98x in Q20 bases; sum-of-contigs Quality
coverage: 17.79x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES

Location/Qualifiers
1..267628
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-10M12"
/clone_lib="RPC1-23"

BASE COUNT

534 a 723 c 757 g 637 t

Query Match 12.0%; Score 200.8; DB 10; Length 2651;
Best Local Similarity 55.8%; Pred. No. 6.6e-40;
Matches 407; Conservative 0; Mismatches 317; Indels 6; Gaps 1;

602 tactacaagattgcagctcattacaagtgaggcattgagtcagctgtttcacaaagcataat 661
Db TACTACAAGATTGCCAGGCATCCGCTGGGCACCTAGGCAGACTCTTCAACAAGTTCAAG 908
662 tttagccgtgttatactatacagaagatgatatggaatgtcgtgctgatttttgactat 721
Db TTCCCGGCCGCTGTGTGTAGTGAGGATGATCTGGAAAGTGCCACAGACTTCTTTGAGTAC 968
722 tttagagctggagctactcttctgacagagacaaagctgattgctattgtcttcttgg 781
Db TTCCAGGCCCATACCCACTGCTGAGAACAGACCCCTCCCTTTGGTGTGTGCTGTGCTGG 1028
782 aatgacaatggacaaggcaggttcgtccca-----agatcctgactcctttaccgctca 835
Db AATGACAAATGGCAAGGACGAGATGGTAGACTCAAGCAAAACCTGAGCTGTCTATCGAACA 1088
836 gactttttctgctgtgattgattgctttcaaaatcaactggtccgaactatctcca 895
Db GACTTTTCTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1148
896 aatggccaaagcttactgggactgactggctgaagctgaaagaaaaatcacagagatcga 955
Db AAGTGGCCCAAGGCCCTTTTGGGACGACTGGATGCGCCGACCTGAGCAGCGGAAAGACCG 1208
956 caattattgcccagaagatttgcagaacgtacaaattttgggagcatgatttctagtttg 1015
Db GCTTGTATTCGTCCAGAAATTTCAAGAACTATGACCTTTTGGTCGCAAGGGGTGTGAGCCAC 1268
956 gggcagtttttaagcagtatctttagaccaatttaagcattaaatgatgtccagggttgattgg 1075

OM of: US-09-591-466c-2 to: EST:* out_format : pfs

Date: Aug 13, 2002 10:41 PM

About: Results were produced by the GenCore software, version 4.5,

Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame*_p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09591466/runat_12082002_095312_4249/app_query.fasta_1.511
-DB=EST -Qfmt=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -XCAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-ALIGN=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-LIST=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09591466 -CGN1_1_4902
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-591-466c-2

Query length: 446

Database: EST:*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 1757.600000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est2:BG594788	+	1143.00	2330.95	768	! BG594788 EST493466 cSTS Solanum
gb_est1:AI486771	+	1116.00	2277.34	643	! AI486771 EST245093 tomato ovary
gb_est2:BG592816	+	1113.00	2271.18	642	! BG592816 EST491494 cSTS Solanum
gb_est2:BI179808	+	1112.00	2267.75	724	! BI179808 EST520753 cSTE Solanum
gb_est1:AI725875	+	1020.00	2078.90	681	! AI725875 BNHG113374 Six-day CG
gb_est2:BG889872	+	1012.00	2062.76	661	! BG889872 EST515723 cSTD Solanum
gb_est1:AW931180	+	988.00	2014.91	574	! AW931180 EST357023 tomato fruit
gb_est1:AW979500	+	974.00	1984.26	673	! AW979500 EST310521 tomato root
gb_est1:AW735827	+	954.00	1943.29	659	! AW735827 EST336595 tomato flower
gb_est2:BE2426174	+	953.00	1941.76	629	! BE2426174 WHE03229_H01_O012s Whea
gb_est2:BE516817	+	859.00	1748.57	603	! BE516817 WHE6620_D08_H162a Whea
gb_est2:BG409413	+	832.00	1693.17	591	! BG409413 00785 leafy spurge Lan
gb_est2:BE805530	+	825.00	1677.43	664	! BE805530 ss47a01.y1 Gm-cl061 G
gb_est2:BG726055	+	764.50	1555.86	506	! BG726055 sae06c11.y1 Gm-cl055 G
gb_hic:AK004760	+	732.00	1469.97	2681	! AK004760 Mus musculus adult ma
gb_est2:BJ170544	+	724.00	1458.72	701	! BJ170544 BJ170544 full length G
gb_est2:BG508571	+	683.00	1388.36	488	! BG508571 sac74e07.y1 Gm-cl072 G
gb_est2:BF066219	+	669.50	1362.30	418	! BF066219 sc090606.y1 Gm-cl065 G
gb_est2:BF177681	+	662.00	1344.26	525	! BF177681 saj64c05.y1 Gm-cl072 G
gb_est2:BF145860	+	642.00	1300.90	635	! BF145860 WHE1833_D12_G232s Seca
gb_est2:BG680531	+	631.00	1274.35	894	! BG680531 602628359f1 NCI_CGAP_S
gb_hic:AF289558	+	608.50	1216.04	2561	! AF289558 Homo sapiens clone p
gb_est1:AW618699	+	608.00	1232.90	530	! AW618699 EST320685 L. pennellii
gb_est2:BG051233	+	601.00	1219.79	472	! BG051233 FMI_57_H04.bl_A003 Fli
gb_est2:C96708	+	586.00	1183.54	758	! C96708 C96708 Rice callus Orzya
gb_est2:BG744010	+	572.50	1155.11	798	! BG744010 602722785f1 NIH_MGC_10
gb_est2:BM137146	+	561.50	1136.31	588	! BM137146 WHE1208_F08_L162s Whea
gb_est2:BE404993	+	561.00	1134.89	544	! BE404993 WHE1208_F08_L162s Whea
gb_est2:BG131020	+	561.00	1134.62	602	! BG131020 EST463912 tomato crown
gb_est2:BI1152075	+	557.50	1124.16	801	! BI1152075 602916527f1 NCI_CGAP_I
gb_est2:BG68233	+	548.50	1104.35	896	! BG68233 6029509817f1 NIH_MGC_15
gb_est1:AW670629	+	539.00	1099.40	546	! AW670629 114751 MARC IPIG Sus s
gb_est2:BG917557	+	531.00	1089.26	823	! BG917557 602820523f1 NCI_CGAP_M
gb_est1:AL503600	+	529.50	1068.17	690	! AL503600 AL503600 Hordeum vulg
gb_est2:BE412855	+	529.50	1068.17	690	! BE412855 MCG007_F12R990625 ITEC
gb_est2:BI415357	+	529.00	1065.29	889	! BI415357 602988551f1 NCI_CGAP_I
gb_est2:BG680449	+	528.00	1064.22	744	! BG680449 602626265f1 NCI_CGAP_S
gb_est2:BE289472	+	517.00	1042.63	677	! BE289472 601087930f1 NCI_CGAP_M
gb_est2:BG481478	+	515.00	1035.81	859	! BG481478 602528861f1 NIH_MGC_21
gb_est1:AI897249	+	507.00	1024.85	528	! AI897249 EST266608 tomato ovary

gb_est2:BM427482 + 506.00 1021.02 1.3e-47 617 ! BM427482 pgf2n.pk006.ol8 Nor
gb_est2:BE821775 + 496.00 999.13 2.1e-46 691 ! BE821775 GM700015A20F4 Gm-r1
gb_est1:AU179654 + 495.00 1000.75 1.7e-46 500 ! AU179654 AU179654 Medaka liv
gb_est2:BG888733 + 484.00 978.36 3.0e-45 488 ! BG888733 EST514584 cSTD Sola
gb_est2:BE382846 + 479.00 963.72 2.0e-44 715 ! BE382846 601297718f1 NIH_MGC

seq_name: gb_est2:BG594788

seq_documentation_block:

LOCUS BG594788 768 bp mRNA linear EST 12-APR-2001
DEFINITION EST493466 cSTS Solanum tuberosum cDNA clone cSTS8020 5' sequence,
mRNA sequence.

ACCESSION BG594788

VERSION BG594788.1 GI:13612928

KEYWORDS EST.

SOURCE potato.

ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 768)

AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemingo,A.,

Bougrin,O., Buell,C.R., Romning,C., Tanksley,S. and Baker,B.

Generations of ESTs from sprouting potato eyes

Unpublished (2000)

CONTACT: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13F-R.

Location/Qualifiers

source

1..768

/organism="Solanum tuberosum"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="cSTS8020"

/clone.lib="cSTS"

/tissue_type="sprouting eyes from tubers"

/dev_stage="12-14 weeks post harvest"

/lab_host="SOLR"

/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; Various sizes of sprouting eyes (2mm to 15mm) were

taken from tubers. The tubers were incubated at 26C in the

dark for 2-3 weeks prior to sprouting. The eyes were

frozen in liquid nitrogen immediately upon removal from

tubers."

BASE COUNT 220 a 149 c 181 g 218 t

ORIGIN

alignment_scores:

Quality: 1143.00 Length: 231

Ratio: 4.991 Gaps: 0

Percent Similarity: 99.134 Percent Identity: 97.403

alignment_block:

US-09-591-466c-2 x BG594788 ..

Align seg 1/1 to: BG594788 from: 1 to: 768

1 MetAtgGLyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17

|||||

74 ATGAGAGGGAACAGTTTTCGTTGATTTACGGTACCTCTCTCGTCGTGGC 123

|||||

17 aLaTgALaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34

|||||

124 TGTCTCGCGCTTCATCTACATACAGATCGGCTTTTCGCGACACAGTCAG 173

|||||

34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHIsCysThr 50

|||||

174 AATATGTAGACCGCTTCGTGTCGAATTAAGACGAGAAATCATTTGACA 223

|||||

51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67

```
|||||
224 AGTCACACAGAGATTGCTTATTGACAAAGATTAGCCAGCAGCAAGGAAGAGT 273
|||||
67 lValAlaLeuGluGluInMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
274 AGTAGCTCTTGAGACAAATGACCGCTCAGGACAGGAGTCCGCCAAT 323
|||||
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
324 TAAGGCTCTTGTTCCAGGATCTTGAAGTAAGGGCATATAAAAGTTAATC 373
|||||
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
374 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTGCTTGGCTGGCAGTCG 423
|||||
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
424 TGCTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATACCAACAT 473
|||||
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
474 CTGTTGCATCAAAATATCCTCTTTTCATATCCAGGATGGATCAATCCT 523
|||||
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
524 GATGTAAAGAAAGCTGCTTTGAGCTATGATCAGCTGACGTATATGCAGCA 573
|||||
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaI 184
|||||
574 CTGGATTTTGAACCTGTGCATACCTGAAAGACCCAGGGAACTGGTTGCAT 623
|||||
184 yrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHis 200
|||||
624 ACTACAAGATTCACCTCATTAACAAGTGGGCATTTGGATCAGCTGTTTTCAC 673
|||||
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
674 AAGCATAATTTAGCCGTGTTATCATACATAGAGAATGATGATGAAATTCG 723
|||||
217 aAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeu 231
|||||
724 TGCTGATTTTGTGACTATTTTGGGGCTGAGCTACTCTCTCTT 766
```

seq_name: gb_est1:A1486771

seq_documentation_block:

LOCUS A1486771 643 bp mRNA linear EST 18-MAY-2001
DEFINITION EST245093 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
cLED11D20, mRNA sequence.

ACCESSION A1486771

VERSION A1486771.1 GI:4382142

KEYWORDS EST.

SOURCE Tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

1 (bases 1 to 643)

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
, Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
,S.D. and Giovannoni,J.

Generation of ESTs from tomato carpel tissue

Unpublished (1999)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>.

Location/Qualifiers

1..643

/organism="Lycopersicon esculentum"

FEATURES

source

```
/cultivar="TA496"  
/db_xref="taxon:4081"  
/clone="cLED11D20"  
/clone_lib="tomato ovary, TAMU"  
/tissue_type="carpel"  
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"  
/lab_host="XLI-Blue MRF"  
/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:  
XhoI; cLED - Tomato Carpel EST Library. OligodT-primed and  
directionally cloned cDNA in vector lamda ZAP II with 5',  
and 3' ends located at the EcoRI and XhoI sites,  
respectively."
```

BASE COUNT 179 a 128 c 144 g 192 t
ORIGIN

alignment_scores:

Quality: 1116.00 Length: 214
Ratio: 5.264 Gaps: 0

Percent Similarity: 99.065 Percent Identity: 96.729

alignment_block:

US-09-591-466C-2 x A1486771 ..

Align seg 1/1 to: A1486771 from: 1 to: 643

99 LeuIleGlyAspValGlnMetProValAlaAlaValValMetAlaCy 115

2 TTAATCGGAATGTCAGATGCCAGTGGCAGCTGTAGTTGTTATGCTTG 51

115 sSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG 132

52 CAGTCGTTCTGACTACCTGGAGAAGACTATAAAATCCATCTTAAAAATACC 101

132 InThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySer 148

102 AAACATCTCTGTCATCAAAATATCCCTTTTCATATCCAGGATGGATCG 151

149 AsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMe 165

152 AATCCTGATGTTAGAAAGCTGCTTTGAGCTATGATCAACTGACGTATAT 201

165 tGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182

202 GCAGCAGCTTGGATTTTGAACCTGTGCATACTGAAAGACCCAGGGAACTGG 251

182 aAlaIatTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeu 198

252 TTGCATACCTACAAGATTGCACGTCATTACAAGTGGCATTTGGATCAGCTG 301

199 PheHisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGl 215

302 TTTACACAGCAATAATTTTACCGCTGTATCATCTAGAAAGATGATATGGA 351

215 utIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuA 232

352 AATTGGCGCTGATTTTGTGACTACTTTTGGGCTGGAGCTACTCTTCTTTG 401

232 sPAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGln 248

402 ACAGAGACAAGTCGATTATGGCTATTTCATCTTGGATCAGCAATGGACAA 451

249 ArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePhePr 265

452 AGGCAGTTTCGTCCTCAAGATCCTTATGCTCTTTACCGCTCAGACTTTTTC 501

265 ocLylLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProL 282

502 TGGTCTTGATGGATGCTTTTCAAAATCAACTTGGTCCGGAACATATCTCAA 551

282 ysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHis 298

552 AGTGGCCGAAGGCTTACTGCGGATGACTGGCTGAGGCTCAAGAAATACAC 601

299 ArgGlyArgGlnPheIleArgProGluValCysArgThrTyr 312
|||||
602 AGAGGTCGACAAATTTATCCCGCAGAAGTTTCAGAACGTTTC 643

seq_name: gb_est2:BG592816

seq_documentation_block:
LOCUS BG592816 642 bp mRNA linear EST 12-APR-2001
DEFINITION EST491494 cSTS Solanum tuberosum cDNA clone cSTS2B9 5' sequence,

ACCESSION BG592816
VERSION BG592816.1 GI:13610956
KEYWORDS EST.
SOURCE potato.

ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 642)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Chiemiango,A.,
Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.

TITLE Generations of ESTs from sprouting potato eyes
JOURNAL Unpublished (2001)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.

FEATURES
source
1..642
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone_lib="cSTS2B9"
/tissue_type="sprouting eyes from tubers"
/dev_stage="12-14 weeks post harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
taken from tubers. The tubers were incubated at 26C in the
dark for 2-3 weeks prior to sprouting. The eyes were
frozen in liquid nitrogen immediately upon removal from
tubers."

BASE COUNT 180 a 124 c 145 g 193 t
ORIGIN

alignment_scores:
Quality: 1113.00 Length: 213
Ratio: 5.275 Gaps: 0
Percent Similarity: 99.061 Percent Identity: 97.183

alignment_block:

US-09-591-466c-2 x BG592816 ..

Align seg 1/1 to: BG592816 from: 1 to: 642

99 LeuIleGlyAspValGlnMetProValAlaValValMetAlaCy 115
|||||
3 TTAATCGGAGATGTCAGATGCCAGTCGAGCTGTAGTTGTATGGCTTG 52
115 sSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG 132
|||||
53 CACTCGTCTGACTACCTGGAGAGACTATTAAATCCATCTTAAATACC 102
132 InThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySer 148
|||||
103 AAACATCTGTCATCAAAATATCCTCTTTTCATATCCAGGATGCATCA 152
149 AsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMe 165
|||||

153 AATCCTGATGTAAGAAAGCTTGTCTTGGAGCTATGATCAGCTGACGTATAT 202
165 tGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182
|||||
203 GCAGCACTTGGATTTCGAACCTGTGCATACCTGAAAGACCAGGGGAAC 252
182 aAlaTyrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeu 198
|||||
253 TTGCATACCTACAGATTGCAGTCATTACAAAGTGGCATTTGGATCAGCTG 302
199 PheHisLysHisAsnPheSerArgValIleIleLeuGluAspMetGI 215
|||||
303 TTTTCAACAGCATAAATTTAGCCGCTGTTATCATCTACTAGAACATGATGGA 352
215 ulleAlaAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuA 232
|||||
353 AAATGTCTGTGATTTTTCAGCTATTATTGAGCTGGAGCTACTCTTCTTG 402
232 sPArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGln 248
|||||
403 ACAGACAGCAAGTCGATTATGGCTATTTCCTTGGAAATGACAATGGACAA 452
249 ArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePr 265
|||||
453 AGGCAGTTCGTCACAGATCCTGATGCTCTTTACCGCTCAGACTTTTTC 502
265 oGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerGluLeuSerProL 282
|||||
503 TGTCTCTGGATGGATGCTTTTCAAAATCAACTGTGTCGAACTATCTTCAA 552
282 YStrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHis 298
|||||
553 AGTGGCCAAAGGCTTACTGGGAGCTAGGCTAAGCTGAAAGAAATATCAC 602
299 ArgGlyArgGlnPheIleArgProGluValCysArgThr 311
|||||
603 AGAGGTCGACAAATTTATTCACCCAGAAAGTTTGCAAAACG 641

seq_name: gb_est2:BI179808

seq_documentation_block:

LOCUS BI179808 724 bp mRNA linear EST 09-JUL-2001
DEFINITION EST520753 cSTE Solanum tuberosum cDNA clone cSTE20K20 5' sequence,
mRNA sequence.

ACCESSION BI179808
VERSION BI179808.1 GI:14645619
KEYWORDS EST.
SOURCE potato.

ORGANISM

Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (bases 1 to 724)

AUTHORS van der Hoeven,R., Bezzerides,J., Bachem,C., Visser,R., Cho,J.,
Chiemiango,A., Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and
Baker,B.

TITLE Generation of ESTs from in vitro grown microtubers

JOURNAL Unpublished (2001)

COMMENT Contact: Cathy Ronning

The Institute for Genomic Research

For clone info: please contact Research Genetics, Libraries

Division tel 1-800-711-6195, email cdna@resgen.com

Seq primer: M13F-R.

FEATURES

source

1..724
/organism="Solanum tuberosum"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="cSTE20K20"
/clone_lib="cSTE"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"

/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue supplied by Christian Bachem and Richard
Visser (Department of Plant Breeding, Wageningen
University, The Netherlands). The cSTA libraries will
attempt to capture the induction and initiation/initial
growth of the tuber in an in vitro system as described in
Bachem et al. (Plant Journal, 1996). Small microtubers
develop from axillary buds attached to stem explants when
placed on a high sucrose medium (10%). Visible
morphological changes occur synchronously at day five in
the axillary buds. The first library, cSTA (1-20) consists
of axillary buds harvested on days 1-3. This targets
those genes involved in induction of the microtubers.
The following libraries, cSTA (21-40) and cSTA (41-60),
capture genes involved in tuber initiation and outgrowth.
This library is noted as P3 in Tanksley lab notebooks."

BASE COUNT 215 a 143 c 171 g 195 t
ORIGIN
alignment_scores:
Quality: 1112.00 Length: 218
Ratio: 5.101 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-591-466C-2 x B1179808 ..

Align seg 1/1 to: B1179808 from: 1 to: 724

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
69 ATGAGAGGGAACAAAGTTTCTGCTTGTATTTACGGTACCTTCTCGTGGC 118
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
119 TGCTCTCCCTTCATCTACATACAGATGCGGCTTTTCGGACACAGCTAG 168
34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
169 AATATGTAGACCGCTTGTCTGCTCAATTCAGACAGAAAATCATTTGACA 218
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
219 AGTCAGACCAAGATTGCTTATTCACAAAGATTAGCCAGCAGCAAGAAAGT 268
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGlnCysArgGlnL 84
269 AGTAGCTCTTGAAGAACAAATGAAGCATCAGGACAGGAGTCCGCGCAAT 318
84 euArgAlaLeuValGlnAspLeuSerLysGlyIleLysLysLeuIle 100
319 TAAGGGCTCTTCTCAGGACTCTTGAAGTAAGGGCATAAAGTTAATC 368
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
369 GGAGATGTCAGATGCCATGCCAGCTGAGTGTATTGCTTGCAGTCG 418
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
419 TACTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATACCAACAT 468
134 exValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
469 CTGTTGCATCAAAATATCTCTTTTCATATCCAGGATGGATCAATCCT 518
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
519 GATGTAAGAAAGCTTGTGAGCTATGCTCAGCTGACGTATATGCACA 568
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||

569 CTTGGATTATGAACCTCTGTGCATACTGAAAGACCAGGGGAAGTGTTCAT 618
184 YrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
619 ACTACAAGATTGCAGCTCATTTACAAGTGGGCAATTGGATCAGCTGTTTCC 668
201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217
669 AAGCATAATTTAGCCCGTGTATCATCTACTACAAGATGATATGGAATTC 718
217 aAla 218
719 TGCT 722

seq_name: gb_est1:A1725875

seq_documentation_block:

LOCUS A1725875 681 bp mRNA linear EST 11-JUN-1999
DEFINITION BNLGH113374 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
to ALPHA-1,3-MANNOSYL-GLYCOPROTEIN, mRNA sequence.

ACCESSION A1725875
VERSION A1725875.1 GI:5044727
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum

REFERENCE 1 (bases 1 to 681)
AUTHORS Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
TITLE ESTs from developing cotton fiber
JOURNAL Unpublished (1995)
COMMENT Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@bnl.bnl.gov
Seq primer: f3 primer.
Location/Qualifiers
1..681

FEATURES
source

/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"
/note="Vector: pBluescript II KS+"

BASE COUNT 202 a 129 c 151 g 197 t 2 others
ORIGIN

alignment_scores:

Quality: 1020.00 Length: 226
Ratio: 4.766 Gaps: 0
Percent Similarity: 94.690 Percent Identity: 78.319

alignment_block:

US-09-591-466C-2 x A1725875 ..

Align seg 1/1 to: A1725875 from: 1 to: 681

152 ValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLe 168
1 GTTAAACTAAGGCTTGAGTTATTAAGGAGCTAAGCTATATATGCACCAT 50
168 uAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyr 185
51 AGATTATCATCCGTGTCATACAGACCGGCTGGGAATTGATCGCATACT 100
185 YrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLys 201

```
|||||
101 ACAAGATTGCCCTCACTACAAATGGCATTGGATGAGTTGTTCTACAAG 150
202 HisAsnPheSerArgValIleIleLeuGluAspMetGluIleAlaI 218
|||||
151 CACAATTTTGACCGAGTAATAATCTGTAAGATCATATGGAATTTGCC 200
218 aAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspL 235
|||||
201 TGATTTTTTGATTACTTGAGCAGCTGCTGCCCTTCTCGAAGAGCA 250
235 ySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPhe 251
|||||
251 AGTCAATTATGGCTGTTCTCTCATGCAATGACAATGGCAAAAGCAGTT 300
252 ValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLeuGI 268
|||||
301 GTGTATGACCCATATGCATCTTATCGCTCAGATTTCTTCTGGTCTGG 350
268 yTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProL 285
|||||
351 CTGCATGCTTACTAAATCTGTATGGAATGAGCTATCACCAAAATGGCAA 400
285 yAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
|||||
401 AAGCTTACTGGGACTGCTGTTGAGATTAAAAAATAATCACAAATGGTCA 450
302 GlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisGI 318
|||||
451 CAATTCCTTCGTCCTGAAGTATGCAGAACATATAATTTGGTGAGCATGG 500
318 ySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuA 335
|||||
501 TTCAAGCATGGGCGAGTTTTCGAAAAATACCTTGCACTTATTAAGATGA 550
335 snAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeuGlu 351
|||||
551 ATGACGTCNAGTGGACTGGAAGTCNAGGATTTGAGCTACCTAAGTGA 600
352 AspAsnTyrValLysHisPheGlyAspLeuValLysLysLysProI 368
|||||
601 GAAAAATATGCCCAATACTTTTCAGACATTTTGAAGCGCGCAAAACCTGT 650
368 ehISgIyAlaAspAlaValLeuLysAla 377
: |||||
651 CCTTGGACAGAGATTCTGCTCTTATGGCA 678
```

seq_name: gb_est2:BG889872

```
seq_documentation_block:
LOCUS      BG889872                      661 bp    mRNA    linear    EST 30-MAY-2001
DEFINITION EST515723 cSTD Solanum tuberosum cDNA clone cSTD15H10 5' sequence,
            mRNA sequence.
ACCESSION  BG889872
VERSION    BG889872.1 GI:14266958
KEYWORDS   EST.
SOURCE     potato.
            Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE  1 (bases 1 to 661)
AUTHORS   van der Hoeven,R., Bezzerides,J., Ewing,E., Cho,J., Chiemiango,A.,
            Bougri,O., Buell,C.R., Ronning,C., Tanksley,S. and Baker,B.
TITLE     Generations of ESTs from dormant potato tubers
JOURNAL   Unpublished (2001)
COMMENT   Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: M13F-R.
            Location/Qualifiers
FEATURES   source
            1. .661
```

```
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTD15H10"
/tissue_lib="cSTD"
/tissue_type="dormant tuber"
/dev_stage="one month post-harvest"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; This library targets genes expressed in dormant
tubers. This library was made from sections of dormant
tuber, avoiding the buds and epidermis. Tubers were stored
for one month post-harvest at 40C. The tuber was peeled,
well away from the surface. Then it was chopped into 1-2
mm cubes and immediately frozen in liquid nitrogen. This
library is noted as P4 in Tanksley lab notebooks."

BASE COUNT      196 a      133 c      158 g      174 t
ORIGIN

alignment_scores:
    Quality: 1012.00      Length: 202
    Ratio: 5.035      Gaps: 0
    Percent Similarity: 99.505      Percent Identity: 98.020

alignment_block:
US-09-591-466C-2 x BG889872 ..
Align seg 1/1 to: BG889872 from: 1 to: 661

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrIleLeuValAlaI 17
|||||
55 ATGAGAGGGAACAAGTTTGTCTTGTATTACGGTACCTTCTCGTCGTGGC 104
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
105 TGCTCTCGCCTTCATCTACATACAGATGCGGCTTTTCGCGACACAGTCAG 154
34 lutyValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
155 AATATGTAGACCCGCTTGCTGTCGCAATTGAACGACAAAATAATATTGTA 204
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
205 AGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGGAAGAGT 254
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
255 AGTAGCTCTTGAGAACAAATGAAGCGTCAGGACCAGGAGTGCCGCAAT 304
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
305 TAAGGGCTCTGTTCAAGGATCTTGAAAGTAAGGGCATAAAAAAGTTAATC 354
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
355 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTGTTATGGCTTGCAAGTCG 404
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
405 TGCTGACTACCTGGAGAGAGACTATTAAATCCATCTTAAATAACCAACAT 454
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
455 CTGTTGGCATCAAAATATCTCTTTTCATATCCAGGATGGATCAAAATCCT 504
151 aspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
505 GATGTAAAGAAAGCTTGCTTTGAGCTATGATCAGCTGACGTATATATGCA 554
167 sleuAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaI 184
|||||
555 CTTGGATTTTGAACCTGTGCATCTACTGAAAGACAGGAGGGAACCTGTTGCAT 604
```

184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
 |||||
 605 ACTACAGATTGCACGTGCTATTACAAGTGGGCATTTGGATCAGCTGTTTCAC 654

201 LysHis 202
 |||||

655 AAGCAT 660

seq_name: gb_est1:AW931180

seq_documentation_block:

LOCUS AW931180 574 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST357023 tomato fruit mature green, TAMU Lycopersicon esculentum
 CDNA clone cLEF43B14 5', mRNA sequence.

ACCESSION AW931180

VERSION AW931180.1 GI:8106581

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE 1 (bases 1 to 574)

AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
 Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.,
 Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato fruit tissue

Unpublished (1999)

CONTACT: CUGI

Clemson University

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

1..574

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEF43B14"

/clone_lib="tomato fruit mature green, TAMU"

/tissue_type="fruit pericarp"

/dev_stage="mature green (3-5 days pre-ripening)"

/lab_host="SOLR"

/note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cLEF - Fruit were tagged at the lcm stage and
 harvested 3-5 days prior to ripening. Fruit were cut in
 half to verify the seeds were indeed 'immature' and the
 seeds and locules were discarded prior to freezing the
 pericarp"

BASE COUNT 154 a 116 c 130 g 173 t 1 others

ORIGIN

alignment_scores:

Quality: 988.00 Length: 190

Ratio: 5.283 Gaps: 0

Percent Similarity: 98.421 Percent Identity: 96.842

alignment_block:

US-09-591-466c-2 x AW931180 ..

Align seg 1/1 to: AW931180 from: 1 to: 574

104 GlnMetProValAlaAlaValValMetAlaCysSerArgThrAspTy 120

|||||

4 CAGATGCCAGTGGCAGCTGTATTGTTATGGCTGCAGTCGTTCTGACTA 53

120 rleuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaAs 137

|||||

54 CCTGGAGAGACGATATAAAATCATCTTAAAAATACCAACATCTGTTCAT 103

137 erLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArg 153
 |||||
 104 CAAAATATCTCTTTTCATATCCAGGATGGATCGAATCCTCATGTAGA 153
 |||||
 154 LysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTy 170
 |||||
 154 AAGCTTGCCTTTGAGCTATGATCAACTGACGTATATGCAGCACTTGGATT 203
 |||||
 170 rGluProValHisThrGluArgProGlyGluLeuValAlaTyrTyrLysI 187
 :|||||
 204 TGAACCTGTGCATACCTGAAAGACCAGGGGAACCTGGTTGCATACCAAGA 253
 |||||
 187 leAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsn 203
 |||||
 254 TTGCACGCTCATTAACAAGTGGCATTGGATCAGCTGTTTCAACAGCATAAT 303
 |||||
 204 PheSerArgValIleIleLeuGluAspAspMetGluIleAlaAlaAspPh 220
 |||||
 304 TTTAGCCGTGTTATCATACTAGAGATGATAGGAAATTCGGGCTGATTT 353
 |||||
 220 ePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysSerI 237
 |||||
 354 TTTTGACTACTTTGAGCTGGAGCTACTCTTCTTGACAGAGACAAGTCGA 403
 |||||
 237 leMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253
 |||||
 404 TTATGGCTATTTCATCTTGAATGACAATGGACAAAGGCAGTTCGTCCAA 453
 |||||
 254 AspProAspAlaLeuTyrArgSerAspPhePheProGlyLeuGlyTrpMe 270
 |||||
 454 GATCCCTTATGCTCTTTACCCTCAGACTTTNTTCCCTGGCTTGGATGGAT 503
 |||||
 270 tLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLysAlaT 287
 |||||
 504 GCITTTCAAAATCAACTTGGTCCGAACTATCTCCAAAGTGGCCGAAGGCTT 553
 |||||

seq_name: gb_est1:AW979500

seq_documentation_block:

LOCUS AW979500 673 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST310521 tomato root deficiency, Cornell University Lycopersicon
 esculentum cDNA clone cLEW209 5', mRNA sequence.

ACCESSION AW979500

VERSION AW979500

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

REFERENCE 1 (bases 1 to 673)

AUTHORS

van der Hoeven,R.S., Garvin,D.F., Matern,A.L., Holt,I.E., Liang,F.,
 Upton,J., Hansen,T.S., Ronning,C.M., Craven,M.B., Bowman,C.L.,
 Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
 and Tanksley,S.D.
 Generation of ESTs from tomato nutrient-deficient roots
 Unpublished (1999)

CONTACT: CUGI

Clemson University

Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

Location/Qualifiers

1..673

/organism="Lycopersicon esculentum"

/cultivar="TA496"

```
/db_xref="taxon:4081"
/clone="cLEW209"
/clone_lib="tomato root deficiency, Cornell University"
/tissue_type="roots"
/dev_stage="5-6 weeks old"
/notes="vector: pBluescript SK-; Site_1: 5' EcoRI; Site_2:
3' XhoI; supplier: Tanksley; Tissue supplied by Dave
Garvin (USDA-ARS, Ithaca, NY 14850). Roots were harvested
from plants grown under the following
deficiencies/stresses: 10 mM Al, Zn, P, K, Fe, N), and
mRNA was isolated from individual treatments. Proportional
aliquots of mRNA of each treatment were mixed and used for
library construction."
```

BASE COUNT 198 a 136 c 159 g 180 t
ORIGIN

alignment_scores:
Quality: 974.00 Length: 201
Ratio: 4.918 Gaps: 0
Percent Similarity: 98.507 Percent Identity: 95.025

alignment_block:
US-09-591-466c-2 x AW979500 ..

Align seg 1/1 to: AW979500 from: 1 to: 673

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
70 ATGAGAGGAACAGATTTTGCTTTGATTACGGTACCTTCGTCGTGGC 119
17 aAlaLeuAlaPheLeuTyrLeuMetArgLeuPheAlaThrGlnSerG 34
|||||
120 TGCTCTCGCCTTCATCTACATACAGATCGCGCTTTTCGTGACACAGTCAG 169
34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
170 AATATGACAGCCGCTTCTGCTGCGATTGAAGCAGAGAAATCATTTGACA 219
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
220 AGTCAGACAGTCTGCTTATTGACAAGATTAGCCAGCAGCAAGGAAGAGT 269
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
270 AGTGGCTCTTGAAGAACAAATGAACGCCAGCAGCAGGAGTCCGACAAAT 319
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
320 TAAGGCTCTTGTTCAGGATCTTGAAGTAAAGGCGATAAAAAAGTTAATC 369
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
370 GGAATGTGCAGATGCCATGGCAGCTGAGTTGTTATGGCTTGCAGTCG 419
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
420 TTCTGACTACCTCGAGAAGACTATAAAATCCATCTTAAATACCAACAT 469
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
470 CTGTTGCATCAAAATATCTCTTTTCATATCCAGATGGATCGAATCCT 519
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
520 GATGTTAGAAAGCTTCTGTTGAGTATGATCAACTGACGTATATGCAGCA 569
167 sLeuAspTyrGluProValHisThrCluArgProGlyGluLeuValAlaI 184
|||||
570 CTTGGATTTTGAACCTGTGCATCTACTGAAAGACCAAGGGAACCTGGTTCAT 619
184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
|||||
```

```
620 ACTACAAGATTGCACGTCATTACAAGTGGCGATTGGATCAGCTGGTTTCC 669
201 Lys 201
||||
670 AAG 672
```

seq_name: gb_est1:AW735827

seq_documentation_block:
LOCUS AW735827 659 bp mRNA linear EST 18-MAY-2001
DEFINITION EST336595 tomato flower buds 0-3 mm, Cornell University
LYCopersicon esculentum cDNA clone cTOA5111 5', mRNA sequence.

ACCESSION AW735827
VERSION AW735827.1 GI:7642666
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 659)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
.F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,
Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue, 0-3 mm buds
Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1..659
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOA5111"
/clone_lib="tomato flower buds 0-3 mm, Cornell University"
/tissue_type="flower"
/dev_stage="0-3mm buds"
/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

BASE COUNT 195 a 132 c 156 g 176 t
ORIGIN

alignment_scores:
Quality: 954.00 Length: 196
Ratio: 4.918 Gaps: 0
Percent Similarity: 98.980 Percent Identity: 95.408

alignment_block:
US-09-591-466c-2 x AW735827 ..

Align seg 1/1 to: AW735827 from: 1 to: 659

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
70 ATGAGAGGAACAAAGTTTGTCTTTGATTACGGTACCTTCGTCGTGGC 119
17 aAlaLeuAlaPheLeuTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
120 TGCTCTCGCCTTCATCTACATACAGATCGCGCTTTTCGTGACACAGTCAG 169
34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
170 AATATGACAGACCGCTTCTGCTGCGATTGAAGCAGAGAAATCATTTGACA 219
```

```
/clone_lib="wheat unstressed seedling shoot cDNA library"
/tissue_type="Etiolated shoot"
/dev_stage="Five day old seedling"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
, germinated and grown aseptically in the dark at room
temperature on filter paper with water, nystatin and
cefotaxime in covered crystallization dishes. Shoots were
harvested. The tissue, total RNA, and poly(A) RNA were
prepared. A cDNA library was made, and the cDNA clones
were in vivo excised to give pBluescript phagemids in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
```

BASE COUNT 195 a 115 c 147 g 172 t
ORIGIN

alignment_scores: Quality: 953.00 Length: 209
Ratio: 4.938 Gaps: 0
Percent Similarity: 92.344 Percent Identity: 81.818

alignment_block:
US-09-591-466C-2 x BE426174 ..

Align seg 1/1 to: BE426174 from: 1 to: 629

135 ValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAs 151
|||||
2 GTTGCTTCAAGTTCCACTATTATATACACAGGATGGAACAAATGGAGA 51

151 pValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisL 168
|||||
52 AGTAAAAAGAAAGCTTGGATTACACTCAATAATACATTTATGCGACATG 101

168 euAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyr 184
|||||
102 TGGATCTTGACCTGTGCGCACTGAAAGACAGGAGAAAAACGTTGCATAT 151

185 TyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPheHisL 201
|||||
152 TACAAGATAGCTAACCACTATAAATGGGCTTGGAGTATTCATTAA 201

201 sHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAlaA 218
|||||
202 GCATGATTTTCGTCGAGTATCATCTTGGAAAGATGACATGGAGATCGCC 251

218 laAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAsp 234
|||||
252 CAGATTTCTTTGACTACTTTGAGCGTGCAGCGAAATTTACTTGACACTGAC 301

235 LysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPh 251
|||||
302 AAGTCGATATGGCTGTTCTTCTTGGCAATGACAATGGCAAAAGCAGTT 351

251 eValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLeuG 268
|||||
352 TGTATATGACCCAAAGGCTCTTTACCGTTGCGATTTCTTCTCGGCGCTTG 401

268 lyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpPro 284
|||||
402 GATGATGCTTACGAGTCAACATGGATGGAGCTGTCCACCAAGTGGCC 451

285 LysAlaTyrTrpAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
|||||
452 AAAGCTTATGGGATGACTGGGTAAAGGAGGTACACAGAGATCG 501

301 gClnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisG 318
|||||
502 GCAGTTTATTCGCCCAAGAGTATGCAGAAACATACAACTTTTGGCGAGCATG 551

```
51 SerGlnThrArgLeuLeuLeuAspLysIleSerGlnGlnGlnGlnArgVa 67  
|||||  
220 AGTCAGACTAGCTGCTTATTGACAAGATTATCCACAGCAGCAAGGAGAGT 269

67 lValAlaLeuGluGlnMetIlyshisGlnAspGlnGlnCysArgGlnL 84  
|||||  
270 AGTGGCTCTTCAAGACAAATGAAGCGCCAGGAGCAGGAGTCCGACAA 319



84 euArgAlaLeuValGlnAspLeuSerLysGlyIleLysLysLeuL 100  
|||||  
320 TAAGGGCTCTTGTTCAGAGTCTTGAAGTAAGGGCATATAAAGTTAATC 369



101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117  
|||||  
370 GGAATGTGCGAGATGCCAGTGGCAGCTGTAGTTGTTATGCTTGCAGTCG 419



117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134  
|||||  
420 TTCTGACTACCTGGAGAGACTATAAAATCCATCTTAAATAACCAACAT 469



134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150  
|||||  
470 CTGTTGCATCAAAATATCTCTTTTCATATCCCAAGGATGATGCAATCCT 519



151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167  
|||||  
520 GATGTTAGAAAGCTTGTCTTGGCTATGATCACTGACGATGATATGCAGCA 569



167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184  
|||||  
570 CTGCGATTTTGACCTGTGCATCTGAAAGACAGGGAAGTGGTGCAT 619



184 yTyrLysIleAlaArgHisTyrLysTyrAlaLeuAsp 196  
|||||  
620 ACTACAAGAATGCACGTCTATTACAAGTGGGATTTGGAT 657



seq_name: gb_est2:BE426174



seq_documentation_block:  
BE426174 629 bp mRNA linear EST 24-JUL-2000  
LOCUS WHE0329_H01_00125 Wheat unstressed seedling shoot cDNA library  
DEFINITION Triticum aestivum cDNA clone WHE0329_H01_001, mRNA sequence.  
ACCESSION BE426174  
VERSION BE426174.1 GI:9424017  
KEYWORDS EST.  
SOURCE bread wheat.  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae  
; Triticeae; Triticum.  
1 (bases 1 to 629)  
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han  
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,  
Seaton,C.L. and Tong,J.C.  
The structure and function of the expressed portion of the wheat  
genomes  
Unpublished (2000)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@w.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stragene SK primer.  
Location/Qualifiers  
1..629  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db_xref="taxon:4565"  
/clone="WHE0329_H01_001"


```


318 lySerSerLeuGlnPhePheLysGlnTyrLeuGluProIleLysLeu 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 GATCAAGCATGGACAATCTTTGATCAATCTTGAACCAATCAATTA 601
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
335 AsnAspValGlnValAspTrpLysSer 343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 AATGATGCTCATATTGACTTGAATTC 628

seq_name: gb_est2:BE516817

seq_documentation_block:
LOCUS BE516817 603 bp mRNA linear EST 08-AUG-2000
DEFINITION WHE620_D08_H162A wheat ABA-treated embryo cDNA library Triticum
aestivum cDNA clone WHE620_D08_H16, mRNA sequence.

ACCESSION BE516817 GI:9740847

VERSION BE516817.1

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.

REFERENCE 1 (bases 1 to 603)

Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y.,
Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey
S.D. and Walker-Simmons,M.K.

The structure and function of the expressed portion of the wheat
genomes - ABA-treated embryo library

Unpublished (2000)

Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA

Tel: 5105595773

Fax: 5105595818

Email: oanders@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20

Seq primer: Clontech Matchmaker 3' AD primer.

FEATURES Location/Qualifiers

source

1..603
/organism="Triticum aestivum"
/cultivar="Brevor (soft, white, winter, common wheat)"
/db_xref="taxon:4565"
/clone="WHE620_D08_H16"
/clone_lib="wheat ABA-treated embryo cDNA library"
/tissue_type="Seed embryo"
/dev_stage="Mature dormant
/lab_host="E. coli DH12S"
/note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI;
Embryos were cut from mature, dormant seeds and imbibed in
25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,
for 12 hr at 22 C. The tissue, total RNA, and poly(A)
RNA were prepared by Steven Verhey in M.K.

Walker-Simmons's lab (USDA-ARS, Washington State Univ.,
Pullman, Washington 99164-6420. A cDNA library was made
by Clontech using a combination of random and oligo dt
primers. Library was plated and archived by Russell
Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid
DNA preparations and DNA sequencing were performed in the
OD Anderson lab (all other authors)."

BASE COUNT 147 a 174 c 120 g 162 t

ORIGIN

alignment_scores:

Quality: 859.00 Length: 199

Ratio: 4.853 Gaps: 0

Percent Similarity: 88.945 Percent Identity: 77.387

alignment_block:

US-09-591-466c-2 x BE516817/rev ..

Align seg 1/1 to reverse of: BE516817 from: 1 to: 603

231 LeuAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnG1 247

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

602 CTTGACACTGACAAGACAATAATGGCTGTTTCATCTTGAATGATAATGG 553

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

247 yGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheP 264

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

552 GCAAAAGCAGTTTCGTTTATGACCCCAAGGCTCTTTACCGTTCGGAGTTCT 503

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

264 heProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSer 280

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

502 TTCCGGGGCTTGGATGGATGCTAACGAAGTCAACATGGATGGAGCTGTGG 453

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

281 ProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAs 297

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

452 CCAAAAGTGGCCCCAAAGCTTATTGGGATGACTGGGTGAGGCTAAAGCAGGT 403

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

297 nhisArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnP 314

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

402 ACACAGAGATCGGCAGTATTTCGGCCAGAAATATGCAGGACATACAAC 353

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

314 heGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGlu 330

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

352 TTGGCGAGCATGATCAAGCATGGGACAATTTCTTGATCAATACTTGAAA 303

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

331 ProIleLysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuse 347

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

302 CCTATCAAGTTAAATGATGCTCATATTGACTGGAATTCGAGGACCTGAG 253

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

347 rTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysL 364

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

252 CTACCTCAAGGAGGACAGTATTTTACCAAAATTCGGGAAAGACGTGGCTA 203

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

364 ysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIle 380

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

202 GCGGCACACCTGTGATCGATCGCTTTGTTGAAGCCCAATCTG 153

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

381 AspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspI1 397

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

152 GATGTGGACGTAAAGGATTACGATATGACAATCAGGCGACTTCGAGCGTAT 103

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

397 eAlaArgGlnPheGlyIlePheGluGluTrpLysAspGlyValProArgA 414

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

102 AGCTCGTCAGTTTGAATATTTGAAGAGTGGAAAGGACGGTGTCCACGGG 53

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

414 laAlaTyrLysGlyIleValValPheArgPheGlnThrSerArgArg 429

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

52 CGGCTTACAAAGGCGTGGTGGTTCCGGTACAAAGGTCGAGCGGG 6

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

seq_name: gb_est2:BG409413

seq_documentation_block:

LOCUS BG409413

DEFINITION 00785 leafy spurge Lambda HybridZAP 2.1 two-hybrid vector cDNA

library Euphorbia esula cDNA clone 24AD 5' similar to

glycosyltransferase like protein, mRNA sequence.

ACCESSION BG409413

VERSION BG409413.1 GI:13315758

KEYWORDS EST.

SOURCE leafy spurge.

ORGANISM Euphorbia esula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Malpighiales; Euphorbiaceae; Euphorbia.

REFERENCE 1 (bases 1 to 591)

AUTHORS Anderson,J.V. and Horvath,D.P.

TITLE Identification of mRNAs expressed in underground adventitious buds

of Euphorbia esula (leafy spurge)

JOURNAL Unpublished (2000)

COMMENT Contact: Anderson JV
Plants Science Research

USDA/ARS, Biosciences Research Lab
1605 Albrecht Blvd., PO Box 5674, Fargo, ND 58105, USA
Tel: 701 239 1263
Fax: 701 239 1252
Email: andersj@fargo.ars.usda.gov
Seq primer: PAD5

FEATURES
source
1..591
Location/Qualifiers
/organism="Euphorbia esula"
/db_xref="taxon:3993"
/clone="24AD"
/clone.lib="leafy spurge Lambda HybriZAP 2.1 two-hybrid
vector cDNA library"
/tissue_type="underground adventitious buds"
/dev_stage="3-day induced (decapitated)"
BASE COUNT 189 a 96 c 141 g 163 t 2 others
ORIGIN

alignment_scores:
Quality: 832.00 Length: 184
Ratio: 4.727 Gaps: 1
Percent Similarity: 95.652 Percent Identity: 79.348
alignment_block:
US-09-591-466C-2 x BG09413 ..
Align seg 1/1 to: BG09413 from: 1 to: 591

```

262 AspPheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG1 278
:::||||| |||||:::|||||:::|||||:::|||||:::|||||
8 GAGTGTTCGGGACTTGGATGGATGGTCTACTAGATCAACCTGGATGA 57

278 uLeuSerProLysTrpProLysAlaTyrTrpAspTrpLeuArgLeuL 295
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
58 ACTATCCCTAAATGGCCTAAAGCTTACTGGGATGACTGGCTCAGCCTAA 107

295 ySGLuAsnHisArgGlyArgGlnPheLeuArgProGluValCysArgThr 311
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
108 AAGAAATCATAAAGCGCGCCCAATTTATTCGTCCARAAATCTGCAGGACA 157

312 TyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTy 328
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
158 TATACTTTGGTGACATGGTTCAGTTTGGGGCAGTTTTTCAACACAGTA 207

328 rLeuGluProLysLysLeuAsnAspValGlnValAspTrpLysSerMeta 345
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
208 TCTTGAGCCCAATTAATGAATGAATATTCAGTCAATTTGGAATCAATGC 257

345 sPLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeu 361
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
258 ATTTGAGTACTGATGAGGAGCAGATATCCAAATATTTTCAGTGACATT 307

362 ValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPh 378
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
308 TTAAGACAAAGCTAAAGCTCAGTGGAAATGATGTTGTTCTTAAGGCATA 357

378 eAsnIleAspGlyAspValArgIleGlnTyrArgAspCinLeuAspPheG 395
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
358 TAACATGAAGGTGATGCGTATTCAGTACAAAGACCAATTAATATTG 407

395 luAspIleAlaArgGlnPheGlyIlePheGluGlnTrpLysAspGlyVal 411
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
408 AAGCATTTGCAGCCAGCTTTGGCATTTTGAAGAAATGAAGGATGGGAATA 457

412 ProArgAlaAlaTyrLys.GlyIleValValPheArgPheGlnThrSerA 428
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
458 CCAAGGACAGCATATAAGGGGAATAGTAGTCTCCGGTATCAAAACACAAA 507

428 rGArgValPheLeuValSerProAspSerLeuArgGlnLeuGlyValGlu 444
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
508 GACGTATATCTCTCGTGGGTCGGGATTCCTGAAACAGCTTGGGAATCGAA 557

```

seq_name: gb_est2:BE805530

seq_documentation_block:

LOCUS BE805530 664 bp mRNA linear EST 06-DEC-2001
DEFINITION ss47a01.y1 Gm-cl061 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl061-1825 5' similar to TR:Q9ST97 Q9ST97 ALPHA-1
3-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE
; mRNA sequence.

ACCESSION BE805530
VERSION BE805530.1 GI:10236642
KEYWORDS EST.

SOURCE soybean.

ORGANISM

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE 1 (bases 1 to 664)

AUTHORS

Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
A., Holla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.

TITLE Public Soybean EST Project

JOURNAL

COMMENT

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130

South Memorial Parkway Huntsville, AL 35801 For further information

call: (800)-533-4363 or contact via email: ccu@resgen.com

Insert Length: 1629 Std Error: 0.00

High quality sequence stop: 419.

FEATURES

source

1..664 Location/Qualifiers

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl061-1825"

/clone.lib="Gm-cl061"

/tissue_type="mature flowers of field grown plants"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; The cDNA library was constructed from mRNA isolated

from mature flowers of field grown plants for the cultivar

Raiden. Complementary DNA was synthesized from mRNA using

a primer consisting of a poly(dT) sequence with a XhoI

restriction site. EcoRI adapters were ligated to the

blunt-ended cDNA fragments followed by XhoI digestion. The

cDNA fragments were directionally cloned into the

EcoRI-XhoI restriction site of the pBluescript vector. The

ligated cDNA fragments were transformed into DH10B host

cells (GibcoBRL). This library was constructed in the

laboratory of Dr. Randy Shoemaker."

BASE COUNT 214 a 121 c 138 g 190 t 1 others

ORIGIN

alignment_scores:

Quality: 825.00 Length: 221

Ratio: 4.209 Gaps: 0

Percent Similarity: 88.688 Percent Identity: 70.588

alignment_block:

US-09-591-466C-2 x BE805530 ..

Align seg 1/1 to: BE805530 from: 1 to: 664

44 GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuLeaspLys11 60

293 rgleuLysGluAsnHisArgGlyArgGlnPheIleArgProGluValCys 309
 |||||
 252 GACTTAAGAGAAATCATAAAGGACACAGTATTATCCGCCCAAGTATGC 301
 |||||
 310 ArgThrTyrAsnPheGlyGluHisGly.SerSerLeuGlyGlnPhePheL 326
 |||||
 302 AGAACATATAATTTT.....GGGTTCTAGTTGGGACAGTTTTAA 342
 |||||
 326 ysglnTyrLeuGluProIleLysLeuAsnAspValcInValAspPrpLys 342
 |||||
 343 AGCAATATCTTCAGGCAATCAAGCTGAATGATGTCAAGGTTGATTGGAAA 392
 |||||
 343 SerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGln 359
 |||||
 393 TTAATGGATCTCAGCTATTACTGGAGGATAAATATCTATGCACCTTGC 442
 |||||
 359 yAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuL 376
 |||||
 443 GAATGTTATTAGAAAGCTACACCTGTCTATGGAGCTGACATGGTTCTAA 492
 |||||
 376 ysaLapheAsnIle 380
 |||||
 493 AGCATCTTAATATA 506

seq_name: gb_htc:AK004760

seq_documentation_block:

LOCUS AK004760 2681 bp mRNA linear HTC 19-JAN-2002
 DEFINITION Mus musculus adult male lung cDNA, RIKEN full-length enriched
 library, clone:1200014F20:mannoside acetylglucosaminyltransferase
 1, full insert sequence.

ACCESSION AK004760

VERSION AK004760.1 GI:12836180

KEYWORDS HTC: CAP trapper.

SOURCE Mus musculus (Strain:C57BL/6J) adult male lung cDNA to mRNA,

clone_lib:RIKEN full-length enriched mouse cDNA library

clone:1200014F20.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (sites)

Authors Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2 (sites)

Authors Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3 (sites)

Authors Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,

Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4 (sites)

Authors The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

REFERENCE

AUTHORS

5 (bases 1 to 2681)
 Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
 Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
 Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
 Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,T.,
 Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
 Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
 Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
 Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
 Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
 Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
 Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamana,K.,
 Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
 Hayashizaki,Y.

TITLE

JOURNAL

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
 URL:http://genome-gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)

COMMENT

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5'-GAGAGAGAGCGCGCACTCGAGTCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. Second
 strand cDNA was prepared with the primer adapter of sequence[5',
 GAGAGAGAGAGGATCCAGAGCTCAATTAATTAATTAACCCGCCCCC 3']. cDNA was
 cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
 XhoI. Host: SOLR.

FEATURES

source

Location/Qualifiers

1..2681
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MGI:1904351"
 /db_xref="taxon:10090"
 /clone="1200014F20"
 /sex="male"
 /tissue_type="lung"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 351..1694
 /gene="Mgat1"
 351..1694
 /gene="Mgat1"
 /note="data source:MGD, source key:MGI:96973,
 evidence:ISS
 mannoside acetylglucosaminyltransferase 1
 putative"

gene

CDS

polyA_signal

polyA_site

BASE COUNT 529 a 736 c 776 g 640 t

ORIGIN

alignment_scores:

Quality: 732.00 Length: 349
Ratio: 2.905 Gaps: 11
Percent Similarity: 72.206 Percent Identity: 42.980

alignment_block:

US-09-591-466c-2 x AK004760 ..

Align seg 1/1 to: AK004760 from: 1 to: 2681

```
103 ValGlnMetProValAlaAlaValValMetAlaCysSerArgThrAs 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
666 GTCAGATCCC.....ATCCTGGTCATTGCTGTGACCCGACGAC 706

119 pTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValA 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
707 T...GTCGCGCGTGGCTGGATAAAGTTGTCACATATCGGCCCTCA...G 750

136 laSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspVal 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 CTGAGCGTTTCCCATATATGTCAGTCAGGACTGTGGGCATGAAGAGACA 800

153 ArgLysLeuAlaLeuSerTyrGly...GlnLeuThrTyrMetGlnHisLe 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 GCACAGGTCAATTGCTTCCTATGGCACTGCAGTCACACATCCGCGCAGCC 850

168 uAsp.....TyrGluProValHisThrGluArgProGlyG 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 GGACCTGAGTAACATTGCCGTGCAGCCAGACACCGCAAGTCCAGGGT. 899

180 luLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAsp 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
900 .....TACTACAAGATTGCCAGGCACCTACCGCTGGGCACCTAGCC 938

197 GlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuGluAspAs 213
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
939 CAGATCTTCAACAAGTTCAAGTTCCCGCGCGCTGTGGTAGTGGAGGATGA 988

213 pMetGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrL 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
989 TCTGGAAGTGGCACCAAGACTCTTTGAGTACTTCCAGGCCACCTACCCAC 1038

230 euLeuAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsn 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1039 TGTCTGAGACAGACCCCTCCCTTGTGTGTGTCTGTCTTGAATGACAAAT 1088

247 GlyGlnArgGlnPheValGlnAsp.....ProAspAlaLeuTyrArgSe 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1089 GGCAAGGAGCAGATGTAGACTCAAGCAAACTGAGCTGCTCTATCGAAC 1138

261 rAspPheProGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerG 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1139 AGACTTTTTCCTGGCCTTGGATGGCTGTGTGGCTGATCTGTGGGCAG 1188

278 luLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeu 294
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1189 AGCTGGAGCCCAAGTGGCCCAAGGCCCTTTTGGACGACTGGATGCCCGA 1238

295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1239 CCTGAGCAGCGGAAAGGAGCGGCTTGTATTCTGCCAGAAATTTCAAGAAC 1288

311 rTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnT 328
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1289 TATGACCTTTGGTCGACAGGGTGTGAGCCACGGGCAGTCTTTTGACCAGC 1338

328 yrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMet 344
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1339 ATCTTAAGTTTCATCAAGCTGAACACGAGCTGCTCCCTTACCAGCTTG 1388

345 AspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLe 361
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1389 GACCTGTGCTACCTGCAGCAGGAGGCTCTATGACCGGACTTC..... 1430

361 uValLysLysAlaLysProIleHisGlyAlaAsp.....AlaValLeuL 376
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1431 .....CTGCCCAGGTCTATGTGTCGCCCCACGCTACAGGTGGAGA 1470

376 ysAlaPheAsnIleAsp.....GlyAspValArgIleGlnTyr 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1471 AAGTAAGGACCAATGATCAGAAGGAGCTGGGGAGGTGCGGTACAGTAC 1520

389 ArgAspGlnLeuAspPheGluAspPheAlaArgGlnPheGlyIlePheG 405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1521 ACTAGCAGACAGACTTCAAGGCTTTGCTAAGGCCCTGGGCTGGGTGTCATGGA 1570

405 uGluTrpLysAspGlyValProArgAlaAlaTyrLysGlyIleValValP 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1571 TGACCTCAAGTCTGGTGTCCCAAGGGCTGGCTACCGTGGCATTGTCACTT 1620

422 heArgPheGlnThrSerArgValPheLeuValSerProAspSer 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1621 TCCAGTTTCGG...GGCCGCGTGTGCACCTGGCACCCCCACAGACA 1664
```



```
94 ..... 94
594 CTTCAAGTGCACAGTATTATTATTCATTGCTCATTTATTCGAAAATGA 643
95 ..... GlyIleL 97
644 AGGAAGCCTTCAACCTGCCGTGCAGATCGGAATGGACGACGAGGTTT 693
97 ySLysLeuIleGly..... 101
694 CAATAGTAATGGGAATTCCTACAGTGAAGAGAGAAGTTAAATCTTACCTC 743
102 ..... AspValGlnMe 105
744 ATAGAACTCTTATCCCTTATTGTAATCTGTATCTCGAAGAGAGTT 793
105 tProValAlaAlaValValMetAlaCysSerArgThrAspTrpLeuG 122
794 GGACTGTGTATAGTATCTTCATAGGAGACAGATACATGATTATCTAA 843
122 luArgThrIleLysSerIleLeuLys...TyrGlnThrSerValAlaSer 137
844 ATGGTGTGTAGCAACCTCGGAGAAAGATTTCTTAAAGAAATCAGTCT 893
138 LysTyrProLeuPheIleSerGlnAspGlySer...AsnProAspValAr 153
894 GGCTTGGTGAATAATATCACCTCCTGAAAGCTATTATCTGACCTGAC 943
153 gLysLeuAlaLeuSerTyrGly.....GlnLeuThrT 164
944 GAACCTTAAGGACACATTTGGAGATCTTAAGAAAGAGTAAGATGAGAA 993
164 yrMetGlnHisLeuAspTyr.....GluProValHisThrGluArgPro 178
994 CAAAGCAAAACCTAGATATTATTTCTTAATGATGATGCTCAGGAAAA 1043
179 GlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLe 195
1044 GGC.....ACATACTAC..... 1055
195 uAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuGlu 212
1056 ..... ATCCAGCTTGAAG 1068
212 sPaspMetGluIleAlaAlaAspPheAspTyrPheGluAlaGlyAla 228
1069 ATGATATTATTCTCAACACAGAATTACTTTAACACCATAAAGAATTTGCA 1118
229 ThrLeuLeuAspArgAspLys...SerIleMetAlaIleSerSerTrpAs 244
1119 CTTCAACTTTCTTCGAGGAATGGATGATCTTGAGTTCTCCAGCTGGG 1168
244 nAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgS 261
1169 ATTCAATTGGT.....AAAATGTTTCAAGCACCTGACCCACTCTCGATTG 1212
261 erAspPhePhe.....ProGlyLeuGlyTrpMetLeu 271
1213 TCGAATTTCATATTATGTTCTATAAGGAGAACCCCATCGACTGGCTCTG 1262
272 SerLysSerThrTrpSerGlnLeu...SerProLysTrpProLysAlaTy 287
1263 GACCATATTCTGTGGTCAAACTCTGCAACCCGGAAGAAAGATGCAAAACA 1312
287 rTrpAspTrpLeuArgLeuLysGluAsnHisArgGlyArgGlnPheI 304
1313 CTGTGAT.....CGACAGAAGGCAAAATCTCGAATTCGT.....T 1347
304 leArgProGluValCysArgThrTyrAsnPheGlyGluHisGlySerSer 320
1348 TCAGACCGTCCCTTTTCCAA.....CACGTTGGTCTGCTATCTTCTACTC 1391
```

```
321 LeuGlyGlnPhe.....PheLysGlnTyrLeuGluProfilely 333
1392 ACAGGAAAAAATTCAGAAACTCAGGATAAAGATTACATGAACCATTTACT 1441
333 sLeuAsn.....AspValGlnValAspTrpLys. 342
1442 GCTCAAAATCCATGTAAACCCCTCGACAGAGTATCTACTTCTTTGAAG 1491
343 .....SerMetAspLeuSerTyrLeuLeuGluAsp..... 352
1492 TCTACCAAGGTCACTACACTGGAGAAACATTACATGGGTGAGGACTTCTTC 1541
353 .....AsnTyrValLysHisPheGlyAspLe 361
1542 TGGGCTATAACCCCTAGTAGCTGGAGACTACATC.....CT 1576
361 uValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaP 378
1577 ATTAAATTCGACACCCAGTCAATGTGGAAGTTATTGTTCCCATAGTG 1626
378 heAsnIleAsp.....GlyAspValArgIleGln..... 387
1627 GCAACAGGATCATCCAGGGGATATTCTCTCAACACACACGGGTGGAAGTT 1676
388 .....TyrArgAspGlnLeuAsp..... 393
1677 CTGGCTTTGAAGAGTGAAGGTTTGGACATCAGCAAGAAACCAAGACAA 1726
394 ....PheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrp...L 408
1727 ACGATTAGAAGATGGCTATTTCAGAATAGGAAATTTGAAACGGTGTG 1776
408 ysAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArg.Ph 424
1777 CGAAGAGGATGTGTGATCCAGCCATAAACCCCATTTTCGGCCTTCGACTT 1826
424 eGlnThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnL 441
1827 TCAGTTATTC...ACAATTCCTGCTTTGGGCCATCTTAATGAGATCCA 1873
441 eu 441
1874 TA 1875
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq.US-09-415-522-5

seq_documentation_block:

; Sequence 5, Application US/09415522A

; Patent No. 6291660

; GENERAL INFORMATION:

; APPLICANT: Gaffney, Thomas

; APPLICANT: Wendland, Juergen

; APPLICANT: Philippsen, Peter

; TITLE OF INVENTION: No. 6291660el Fungal Genes Required For No. 6291660mal Growth

; FILE REFERENCE: CGC2046

; CURRENT APPLICATION NUMBER: US/09/415,522A

; CURRENT FILING DATE: 1999-10-08

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 6216

; TYPE: DNA

; ORGANISM: Ashbya gossypii

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(6216)

; US-09-415-522-5

alignment_scores:

Quality: 105.00

Length: 389

Ratio: 0.640

Gaps: 18

Percent Similarity: 42.159 Percent Identity: 19.280

alignment_block:

US-09-591-466C-2 x US-09-415-522-5 ..

Align seg 1/1 to: US-09-415-522-5 from: 1 to: 6216

```
55  LeuLeuIleAspLysIleSerGlnGlnGlnArgValValAlaLeuG1 71
||||| ||||| ||||| ||||| ||||| |||||
1210 CTGTGGACGACCAAGATTTC.....CAGTCCCTGGTGATCTACTCGA 1253

71  uGluGlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuV 88
||||| ||||| ||||| ||||| ||||| |||||
1254 TACGATTTCCTCCAGATGAAGACAGCGCGGAGGAGCTGAAGGTGGCGA 1303

88  aGlnAspLeuGluSerLysGlyIleLysLysLeuIleGlyAspValGln 104
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1304 TACCGAGAAACAGACGGTA...CTGGTCAAGCTGACCGGCTTGCAAAAT 1350

105 MetProValAlaAlaValValMetAlaCysSerArgThrAspTyrLe 121
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1351 GAACCCATCCAGTCC.....GCGAACTCGATCTTT 1382

121  uGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerL 138
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1383 AATAAG...GTGAGAGCTTCTGAAACTTGATACCGAAGAGTTGCCA 1429

138  ysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLys 154
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1430 AC.....CAGATTCCACAG 1443

155 LeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGln.....HisLe 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1444 ATCAATCTAACCTTTAATAGGTATGGAGCCCAACAGCGATTATTCCT 1493

168  uAspTyrGluProValHisThrGluArgProGlyGluLeu..... 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1494 ACTTTACGACTCTCAATATACACAAACCGCTGGAACCTAACCCATTGG 1543

182 .....ValalaTyrTyrLysIleAlaArgHisTyr 191
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1544 TATTTTCAACGATAAAATGTACAGTAT..... 1572

192  LysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValI1 208
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1573 .....TTGAGTCGCTTAATGGTGCTCATATCTTCTCGAA..... 1608

208  eIleLeuGluAspAspMetGluIleAlaAlaAspPhePheAspTyrPheG 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1609 .....G 1609

225  luAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSer 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1610 AGACGGGATTACG.....CCGAAGAAACGAGCGGAGGTTTGACA 1650

242  SerTrpAsnAspGlnArgGlnPheValGlnAspProAspAlaLe 258
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1651 AAATGGTGCCAATTTGGGATGCAAG..... 1674

258  uTyrArgSerAspPheProGlyLeuGlyTyrMetLeuSer..... 272
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1675 .....TTTGAGCGACTTGGGACATGGTCTCATGCTG 1708

273 .....LysSerThr 275
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1709 CAATTGGCAGTAATATGCTCCATCCCGTTTACGCTTGACAGGACG 1758

276  TrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspTrpLe 292
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1759 TGCATATGTGCTGACTTCTACTTGAAGATAAATTTTAAAGGATGGGT 1808

292  uArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluValC 309
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
1809  A.....CCCACGATTG 1819
309  ysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePhe 325
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1820 TCCAGTTGGATCGCAGGCAATGTCTCCAAGTCGATGAACAGTGTTTTC 1869

326  LysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLys 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1870 .....ATACTAGCCCACTAATTAAACGATGCC..... 1899

342  sSerMetAspLeuSerTyrLeuLeuGluAsnTyrValLysHisPheG 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1900 .....TTGTGAGGCAATGTGATCCCTTACTTTG 1930

359  lyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLeu 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1931 GCGACTTAGTC.....ATTCCTCCGATGAT..... 1956

376  LysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnLe 392
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1957 .....CTACCCAGACAGACAGTATAAGTACTTGGAG..... 1989

392  uAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLysA 409
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1990 .....AAAAGATACCCCGCACAAAAATGCCTTTTACAAGTGGCAGC 2032

409  spGlyValProArgAla 414
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2033 AGAGACTAGACGCGCA 2049

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-840-236-2

seq_documentation_block:
; Sequence 2, Application US/08840236
; Patent No. 5922578
; GENERAL INFORMATION:
; APPLICANT: MARUTA, Kazuhiko
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME WHICH
; TITLE OF INVENTION: FORMS NON-REDUCING SACCHARIDE FROM REDUCING AMYLACEOUS
; NUMBER OF INVENTION: SACCHARIDE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840.236
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/505,448
; FILING DATE: 21-JUL-1994
; APPLICATION NUMBER: JP 190183/1994
; FILING DATE: 21-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 189706/1995
; FILING DATE: 04-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: MARUTA-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
```



```

218 aAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspL 235
      |||
1291
235 ysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPhe 251
      ||| :||| :||| :||| :||| :|||
1298 AAATCACACTAAATGCCACTAGCACACATCATACT.....AGGTTT 1338
      :|||
252 ValGlnProAspAlaLeuTyrArgSerAspPheProClyLeuG1 268
      :|||

```

```

3339 AGTCAAGATGTA..... 1350
268 yTpmMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProL 285
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
1351 ..... AGGATGCAAAATAGTGTATTAACTGAATTTCTCGAAGATGG...A 1393
285 yslatYtrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
      | : : : : | : : : : | : : : : | : : : : | : : : : |
3394 AAAATGAAGTCGAGGAATGG.....CAT 1416
302 GlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisG 318
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
417 AGTATCATAAATCCAAAGGTATCAAGAAATGATGAATATAGATATATCA 1466
318 ySerSerLeuGlyGlnPhePhe..... 325
      : : : : : | : : : : | : : : : | : : : : | : : : : |
467 GCTTTTAGTGGGAAGTTTTATGAGGGATCTCTAATGATTTTAAGGAGA 1516
326 .....LysGlnTyr...LeuGluProIleLysLeuAsnAspValGlnVal 339
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
517 GAATAAGCAACATATGATAAAAAGTGTACAGAGAGCTAAGATAAAATACC 1566
340 AspTrpLysSerMetAspLeuSerTyr.....LeuLeu 350
      : : : : : | : : : : | : : : : | : : : : | : : : : |
567 TCATGGAGAAATCAAAATATAAGAAATATCAAAATAGAGTAATGGAATTAGT 1616
350 uGluAspAsnTyrVal...LysHisPheGlyAspLeuValLysLysAla 366
      | : : : : | : : : : | : : : : | : : : : | : : : : |
567 GGAAGAAACTTTACCATAAGGATTTC..... 1644
366 ySProlleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGly 382
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
565 .....ATTAAGAGTTTC..... 1656
383 AspValArgIleGlnTyrArgSpGlnLeuAspPheGluAspIleAlaAr 399
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
557 .....ATGAAATTTGAAAGTAAGATAAG 1679
399 gGlnPheGlyIlePheGluGlu.....TrpL 408
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
580 AAGGATAGGATGATTAAAGAGCTTATCCTTGGTCGCATTAAAAATATATCT 1729
08 yAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPhe 424
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
30 CAGCCGGTATACCT...GATTTTATCAGGGAACAGAAATATGGCGATAT 1776
25 GlnThrSerArgValPheLeuValSerProAspSer 437
      : : : : : | : : : : | : : : : | : : : : | : : : : |
      : : : : : | : : : : | : : : : | : : : : | : : : : |
77 .....TTACTTACAGATCCAGATAC 1797

```

```

seq_documentation_block:
; Sequence 5, Application US/08840236
; Patent No. 5922578
; GENERAL INFORMATION:
; APPLICANT: MARUTA, Kazuhiko
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME WHICH
; TITLE OF INVENTION: FORMS NON-REDUCING SACCHARIDE FROM RE
; TITLE OF INVENTION: SACCHARIDE
; NUMBER OF SEQUENCES: 19

```



```

alignment_block:
US-09-591-466C-2 x US-08-505-448A-2 ..
Align seg 1/1 to: US-08-505-448A-2 from: 1 to: 2160

21 PheIleTyrlleGlnMetArgLeuPheAlaThrGlnSerGluTyrrValas 37
||||| : : : : : |||
826 TTCITAAATTACTCCAACCTTACTGTTTAAATTTAATCAAGAGATAATGGA 875

37 pArgLeuAlaAlaIleGluAlaGluAsnHisCysThrSerGlnThrA 54
| : : : : : ||| : : : : : |||
876 CAGTATATATAGAGATTTTCACGCGGAGAAATATCTCTAAGTGAAGT. 924

54 rgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValaIalalaLeu 70
||| ||| : : : : : |||
925 .....ATRAAGAAATATAAAGCGCAA..... 945

71 GluGluGlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLe 87
946 .....AT 947

87 uValGlnAspLeuGluSerLysGlyIleLysLysLeuIleGly..... 101
: : : : : ||| : : : : : ||| : : : : : |||
948 AATTGATGAGCTATTATTAGTTATGAAGTTAAAGATTAGCATCAACAAGTAG 997

102 .....AspValGln 104
|||||
998 GAATTAGCTACGATATATTGAGAGATTACCTTCTTGTATAGATGCTAC 1047

105 MetProValAlaAlaValValMetAlaCysSerArgThrAspTyrlle 121
||| : : : : : ||| : : : : : ||| : : : : : |||
1048 AGAACTTTATGCTAATCAGATTGTAAAGAGTGTGATAAGACCACATGAGAT 1097

121 uGluArgThrIleLysSerIleLeuLysTyrrGlnThrSerValAlaSerL 138
||||| : : : : : |||
1098 AGAGGAAGCAACCAA..... 1113

138 ySTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLys 154
|||||
1114 .....AGAAATCCAGAGGCTTATACT 1134

155 LeuAlaLeuSerTyrrGlyGlnLeuThrTyrrMetGlnHisLeuAspTyrrG 171
| : : : : : ||| : : : : : ||| : : : : : |||
1135 AAATTACACAATATATGCCAGCAGTATACGCTAAA.....GCTTATGA 1178

171 uProValHisThrGluArgProGlyGluLeuValAlaTyrrLysIleA 188
| : : : : : ||| : : : : : ||| : : : : : |||
1179 AGATACTTTCTCTTTATGATACAAATAGATTAATATCCATAAATGAGGTTG 1228

188 la.....ArgHisTyrrLysTyrrAlaLeuAspGlnLeuPheHisLys 201
| : : : : : ||| : : : : : ||| : : : : : |||
1229 GAAGCGATTACGATATATAAGATATCGCTCATCAG...TTTCATGTA 1275

202 HisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAlaA 218
||| : : : : : |||
1276 TTTAATCAAAAGCA..... 1290

218 aspPhePheAspTyrrPheGluAlaGlyAlaThrLeuLeuAspArgAspL 235
|||
1291 .....AGAGGAA 1297

235 ysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPhe 251
|| : : : : : ||| : : : : : ||| : : : : : |||
1298 AAATCACACATAAATGCCACTAGCACACATCATACT.....AAGTTT 1338

252 ValGlnAspProAspAlaLeuTyrrArgSerAspPhePheProGlyLeuG 268
: : : : : |||
1339 AGTGAAGATGTA..... 1350

268 yTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProL 285
| : : : : : ||| : : : : : ||| : : : : : |||
1351 .....ACATCAAAATAGCTGTAATTAAGCTGAATTTCTCTGAAGATGG...A 1393

```

alignment_scores:	
Quality:	102.50
Ratio:	0.551
Percent Similarity:	40.173
Percent Identity:	17.495
Gaps:	21
Length:	463

```
285 ysAlaTyrTrpAspTrpLeuArgLeuLysGluAsnHisArgGlyArg 301
|| : : : : : |||
1394 AAAATAAGTCGAGGAATGG.....CAT 1416
: : : : : |||
302 GlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHisL 318
: : : : : ||| : : : : : |||
1417 AGTATCATATAATCCAAAGGTATCAAGAAATGATGATATAGATATATCA 1466
: : : : : ||| : : : : : |||
318 ySerSerLeuGlyGlnPhePhe..... 325
: : : : : ||| : : : : : |||
1467 GCTTTAGTGGCAAGTTT...TATGAGGGATTCTCTAATGATTTTAAAGGAGA 1516
: : : : : ||| : : : : : |||
326 .....LysGlnTyr...LeuGluProIleLysLeuAsnAspValGlnVal 339
: : : : : ||| : : : : : |||
1517 GAATAAAGCAACATATGATAAAAGTGTACAGAAAGCTAAGATAAATACC 1566
: : : : : ||| : : : : : |||
340 AspTrpLysSerMetAspLeuSerTyr.....LeuLe 350
: : : : : ||| : : : : : |||
1567 TCATGGAGAAATCAAAATAAAGAAATATGAAATAGATGAATAGTAGT 1616
: : : : : ||| : : : : : |||
350 uGluAspAsnTyrVal...LysHisPheGlyAspLeuValLysLysAlaL 366
: : : : : ||| : : : : : |||
1617 GGAAGAAACTTTTACCAATAAGGATTCT..... 1644
: : : : : ||| : : : : : |||
366 ySProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGly 382
: : : : : ||| : : : : : |||
1645 .....ATTAAAGTTTC..... 1656
: : : : : ||| : : : : : |||
383 AspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaAr 399
: : : : : ||| : : : : : |||
1657 .....ATGAAATTTGAAAGTAAGATAAG 1679
: : : : : ||| : : : : : |||
399 gGlnPheGlyIlePheGluGlu.....TrpL 408
: : : : : ||| : : : : : |||
1680 AAGGATAGGGATGATTAAAGAGCTTATCCTTGTCGATTAAATAATATGT 1729
: : : : : ||| : : : : : |||
408 ySAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPhe 424
: : : : : ||| : : : : : |||
1730 CAGCCGGTATACCT...GATTTTATCAGGGAACAGAAATATGGCGATAT 1776
: : : : : ||| : : : : : |||
425 GlnThrSerArgArgValPheLeuValSerProAspSer 437
: : : : : ||| : : : : : |||
1777 .....TTACTTACAGATCCAGATAAC 1797
: : : : : ||| : : : : : |||
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:us-08-505-448A-5

```
seq_documentation_block:
; Sequence 5, Application us/08505448A
; Patent No. 5976856
; GENERAL INFORMATION:
; APPLICANT: MARUTA, Kazuhiko
; APPLICANT: KUBOTA, Michio
; APPLICANT: SUGIMOTO, Toshiyuki
; TITLE OF INVENTION: RECOMBINANT THERMOSTABLE ENZYME WHICH
; TITLE OF INVENTION: FORMS NON-REDUCING SACCHARIDE FROM REDUCING AMYLACEOUS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W. Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/505,448A
; FILING DATE: 21-JUL-1994
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 21-JUL-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 04-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: MARUTA=3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2160 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-505-448A-5

alignment_scores:
Quality: 102.50 Length: 463
Ratio: 0.551 Gaps: 21
Percent Similarity: 40.173 Percent Identity: 17.495

alignment_block:
US-09-591-466C-2 x US-08-505-448A-5 ..

Align seg 1/1 to: US-08-505-448A-5 from: 1 to: 2160

21 PheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerGluTyrValas 37
||| : : : : : |||
826 TTCTTAAATTTACTCCAACCTTACTGTTTAAATTTTAAATCAAGAGATAATGGA 875
: : : : : ||| : : : : : |||

37 pargLeuAlaAlaIleGluAlaGluAsnHisCysThrSerGlnThra 54
| : : : : ||| : : : : : |||
876 CAGTATATATGAGAAATTTCCACAGCGGAGAAAATATCTATAAGTGAAAGT. 924
: : : : : ||| : : : : : |||

54 rgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValValAlaLeu 70
||| ||| : : : : : |||
925 .....ATAAGAAATATAAAGCGCAA..... 945
: : : : : ||| : : : : : |||

71 GluGluGlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLe 87
: : : : : ||| : : : : : |||
946 .....AT..... 947
: : : : : ||| : : : : : |||

87 uValGlnAspLeuGluSerLysGlyIleLysLysLeuIleGly..... 101
: : : : : ||| : : : : : |||
948 AATTGATGAGCTATTTAGTTATGTAAGTTTAAAGATTAGCATCACAACTAG 997
: : : : : ||| : : : : : |||

102 .....AspValGln 104
: : : : : ||| : : : : : |||

998 GAATTAGTACCATATATTGAGAGATTACCTTTCTTGTATAGATGTGTAC 1047
: : : : : ||| : : : : : |||

105 MetProValAlaAlaValValMetAlaCysSerArgThrAspTyrLe 121
||| : : : : : |||
1048 AGAAGCTTATGCTAATCAGATTGTAAGAGTGTGATAAGCAACCAATGAGAT 1097
: : : : : ||| : : : : : |||

121 uGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerL 138
: : : : : ||| : : : : : |||

1098 AGAGGAAGCAACCAAA..... 1113
: : : : : ||| : : : : : |||

138 ySyrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLys 154
: : : : : ||| : : : : : |||

1114 .....AGAAATCCAGAGGCTTATACT 1134
: : : : : ||| : : : : : |||

155 LeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGl 171
: : : : : ||| : : : : : |||

1135 AAATTACAACAATATATGCCAGCAGTATACGCTAAA.....GCTTATGA 1178
: : : : : ||| : : : : : |||
```



```
396 CTC.....GAGGATGCTCTGTGTCTATGGCTGGC..... 425
184 rTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisL 201
426 .....CGAATCATGCACA 438
201 yshisAsnPheSerArgValIleIleLeuGluAspMetGluIleAla 217
439 AGGCACCTCTGGTTCTAACTCGTC..... 464
218 AlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAs 234
465 .....TTTATGACCTGCACAGCTGGTGGTCAAGGTC..... 497
234 pLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnP 251
498 ....CAGGTATGGCTGATCGCGAGTAATACAGACTTGGATGAGCGTGAAT 543
251 heValGlnAspProAspAlaLeuTyrArgSerAspPhe..... 263
544 TTTCCAAATTCATCTTAATGTGAACGGTGGGACATAGTTGGTATCACT 593
264 ...PheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLe 279
594 GGCTTTCCAGGC.....AAAAGTAAGAAGGGTGAAC 625
279 user..... 280
626 TAGATTTTCCCAAGACTTTGTGTGTCTCATTTGTTGCATATGA 675
281 ..ProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGlu 296
676 TGCCAAAGGCAAAAGTCTGCTGCTGCGGATAATGCAAAATTTGAAGAAA 725
297 Asn..... 297
726 AATCCATGGGTACCGAGGAAGTACCGAGGAATCCTGAAACATATATTTGAA 775
298 .....HisArgGlyArgGln.....PheIleArgP 306
776 AGATCAGGAACCTAGTATCGGTGAACCCATTGGGATTTCATGCTTAACC 825
306 roGluValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGly 322
826 CAGAGGTTCAGAAATATTT.....AAGACCCGGTCTAAATCAIT 866
323 GlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnVa 339
867 TGTACATTAGGAGGTTCCTGTGATGACCTTGATTTCTGGAGGTGGAAC 916
339 lAspTrpLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValL 356
917 ACCAATGATGACATGATTCGTGCGAGCTGCGCCCGTCCATTTGTAA 966
356 yshisPheGlyAsp..... 360
967 CTCATCAATGATCTTACATGAGGTATTTCATGAGGATTCCTCCAGAA 1016
361 LeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLeu.... 375
1017 CTGTATCTTAAGGAGTTGTTGTGTGCTGACCTGGATCGTGTATTATCAAT 1066
376 ....LysAlaPheAsnIleAspGly..... 392
1067 TGTAAACAATTTAGGAATGAGGGCATAGATTTGACCCCAATAATCTCGAGT 1116
383 .....AspValArgIleGlnTyrArgAspGlnLeuAspPheGlu 395
1117 TTACTACCTGTGAGTTCTATATCGCTTATAAGGACTACACGACTTGATC 1166
396 AspIleAlaArgGlnPhe.....GlyIlePheGluGluTrpLysAspGI 410
1167 GATATAACAGACCAATTTGTGATGTTGTTGTTAAGGAA..... 1205
```

```
410 yValProArgAlaAlaTyrLysGlyIleValVal 421
1206 .CTTACCAAGCAGCTATAAAATCAAGTATCATG 1238
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-117-860-23
seq_documentation_block:
; Sequence 23, Application US/09117860A
; Patent No. 6338955
; GENERAL INFORMATION:
; APPLICANT: OGURI, Suguru
; APPLICANT: MINOWA, Mari
; APPLICANT: YOSHIDA, Aruto
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: TAKEUCHI, Makoto
; TITLE OF INVENTION: NOVEL 1-4 N-ACETYLGLUCOSAMINYLTRANSFERASE AND GENE
; FILE REFERENCE: 081356/0119
; CURRENT APPLICATION NUMBER: US/09/117, 860A
; EARLIER APPLICATION NUMBER: WO PCT/JP97/04546
; EARLIER FILING DATE: 1997-12-10
; EARLIER APPLICATION NUMBER: JP 161462/1997
; EARLIER FILING DATE: 1997-06-18
; EARLIER APPLICATION NUMBER: JP 332411/1996
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
; LENGTH: 2115
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (136)..(1740)
US-09-117-860-23
```

```
alignment_scores:
Quality: 98.50 Length: 568
Ratio: 0.402 Gaps: 31
Percent Similarity: 43.134 Percent Identity: 20.423
alignment_block:
US-09-591-466c-2 x US-09-117-860-23 ..
Align seg 1/1 to: US-09-117-860-23 from: 1 to: 2115
10 LeuArgTyrLeuValAlaLeuAlaPheIle.....TyrII 24
142 CTCGCAATGGAACGTGAGCCACTGCTTTAGCATTTATCACTTCCTTCCT 191
24 eGlnMet.....ArgLeuPheA 30
192 TACTTTGTCTGTGTACTACTGCGCAAAATGGAAAGAAAACTGATTG 241
30 laThrGlnSerGlyTrpValAspArgLeuAlaIleGluAlaGlu 46
242 CTTATCAACGAGAATTCCTT...GCTTTGAAAGAACGCTCTTCCAATAGCT 288
47 AsnHisCysThrSerGlnThrArgLeuLeuIleAspLysIleSerGlnGI 63
289 GAACACAGATCTCACAGCGCTCTTCTGATTAATACGATTGTGCAACA 338
63 nGlnGlyArgValAlaIle.....LeuGluGluGlnMetL 75
339 GTTCAACGCTGTAGGAGCAGCAAAACAATGGAAGATGCGTTGAATA 388
75 yshisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAsp... 90
389 AGTTTTCAGATAATACCTAAAGCTGTTAAAGAGGTTAACAAGCAAAAA 438
```

```
91 .....LeuGluSe 93
439 TCTCTCAAGTGCAAGTATTATTATTCATTTCCTCATTTATTGAAAA 488
93 rLysGly.....
489 TGAAGGAAGTCTTCAACCTCTGTACAGATTGCAACGGGAACAGGAG 538
95 .....
539 TTCAATAGTCATGGGCATTCCACAGTGGAAGAGAAAGTTAAATCTTAC 588
96 .....ileLysLysLeuIleGlyAspValG1 104
589 CTCATAGAAACTCTTCATTCCTCTATTGATACCTGTATCCTGAAGAGAA 638
104 nMetProValAlaValValMetAlaCysSerArgThrAspTyrL 121
639 GTTGGACTGTGTATAGTCTTTCATAGGAGACAGATATTGATTATG 688
121 euGluArgThrIleLysSerIleLeuLys...TyrGlnThrSerValAla 136
689 TACATGTGTGTAGCAACCTGGAGAAGAATTTCTTAAGAAATCAGT 738
137 SerLysTyrProLeuPheIleSerGlnAspGlySer...AsnProAspVa 152
739 TCTGGCTTGTGGAGTCATATCACCCCTGAAAGCTATTATCTGCATT 788
152 lArgLysLeuAlaLeuSerTyrGly.....GlnLeuT 163
789 GACAACGATTAAGAGACATTTGGAGACTCCAAAGAAAGAGTAAAGTGA 838
163 hrTyrMetGlnHisLeuAspTyr.....GluProValHisThrGluArg 177
839 GAACAAGCAAAACCTAGATTACTGTTTCTTAATGATGTATGCTCAAGAA 888
178 ProGlyCluLeuValAlaTyrLysIleAlaArgHisTyrLysTrpAl 194
889 AAGGCC.....ATATATTAC.....
194 aLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeuG 211
904 .....ATTCAGCTTG 913
211 luAspAspMetGluIleAlaAlaAspPheAspTyrPheGluAlaGly 227
914 AAGATGATATTATGTCAACAAATAATTTTAAATACCATATAAAAAATTT 963
228 AlaThrLeuLeuAspArgAspLys...SerIleMetAlaIleSerSerTr 243
964 GCATTCAACTTCTCTCGAAGTGGATGATCTAGAGTTCCTCCAGCT 1013
243 pAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrA 260
1014 GGCCTTCATTGGT.....AAAATGTTCAAGCGCGGATCTTACTCTGA 1057
260 rgSerAspPhePhe.....ProGlyLeuGlyTrpMet 270
1058 TTCTAGATTCAATTCATGATGTTTCAAGGAGAAACCCATTGATGGCTC 1107
271 LeuSerLysSerThrTrpSerGluLeu...SerProLysTrpProLysAl 286
1108 CTGGACCATATTCTCTGGTGAAAGCTGCAACCTGAAAGATGCAAA 1157
286 atyTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGlyArgGlnP 303
1158 ACATTGTGAT.....ACACAGAAAGCAAAATCTGCGAATTCGC... 1194
303 heIleArgProGluValCysArgThrTyrAsnPheGlyHisGlySer 319
1195 ..TTCAGACCTCCCTTTTCCA.....CATGTGGTCTGCACATCATCA 1236
320 SerLeuGlyGlnPhe.....PheLysGlnTyrLeuGluProI1 332
```

```
1237 CTATCAGGAAAAATCCAAAAACTCACGATAAAGATTATATGAACCAATT 1286
332 eLysLeuAsn.....AspValGlnValAspTrpL 342
1287 ACTTCTTAAATCCATGTAACCCACCTCGCGAGGTATCTACTTCCTTGA 1336
342 ys.....SerMetAspLeuSerTyrLeuLeuGluAsp... 352
1337 AGGTCTACCAAGGCATACGCTGGAGAAAACCTTACATGGGAGAGGATTC 1386
353 .....AsnTyrValLysHisPheGlyAs 360
1387 TCTGGGCTATCACACGATAGCTGGAGACTACATC..... 1422
360 pLeuValLysLysAlaLysProIle..... 368
1423 .TTGTTTAAATTTGATAAACCACTCAATGTAGAAAGTTATTTGTTCCATA 1471
369 .....HisGlyAlaAspAlaValLeuLysAla..... 377
1472 GCGGCAACCAAGAACATCTCTGGAGATATTCTGCTAAACACAACTGTGAA 1521
378 .....PheAsnIleAspGly...AspValArgIleGlnTyrArgAs 390
1522 GTTTTCCTTTTAAGAGTGAAGGTTTGGAAATAGCAAAAGAAACCAAGA 1571
390 pGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluT 407
1572 CAAA...CGATTAGAAGTGGCTATTTCAGAAATAGCAAAATTTGAGAATG 1618
407 rp...LysAspGlyValProArgAlaAlaTyrLysGlyIleValValPhe 422
1619 GTGTTCCAGAAGGAATGGTGATCCAAAGTCTCAATCCCATTTCCAGCTTT 1668
423 Arg.PheGlnThrSerArgArgValPheLeuValSerProAspSerLeuA 439
1669 CGACTTTTCAGTTATTC...AGAATTCGTGTTTGGGCCATTCTTAATGA 1715
439 rg 439
1716 GA 1717
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-443-501A-2

```
seq_documentation_block:
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; FILE REFERENCE: 30062-20031-00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
```

```

; ORGANISM: Artificial Sequence
; FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

alignment_scores:
    Quality: 97.50      Length: 569
    Ratio: 0.393       Gaps: 29
    Percent Similarity: 43.585    Percent Identity: 19.332

alignment_block:
US-09-591-466C-2 x US-09-443-501A-2  ..

Align seg 1/1 to: US-09-443-501A-2 from: 1 to: 71989

28 LeuPheAlaThrGlnSerGluTy rValAspArgLeuAlaAlaIleG1 44
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7841 CTGTTCCGGCGCGGTCTCAGCAGCTGCCTATGCAGCTCGCCGTGGTGC 7890

44 uAlaGluAsnHisCysThrSerGlnThrArgLeuLeuLeuAspLysIleS 61
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7891 GGCG.....CGCAAGACGGCTCAGTACGTAAGAAGACTTT 7922

61 erGlnGlnGlnGlyArgValAlaLeuGluGlnMetLysHisGln 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7923 CGCGCGGTTCGGCGGCACTTGGCGG.....CGGTGCGCGGACGAG 7963

78 AspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeuGluSerLy 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7964 GGGGCAACGCCCGAACAACATGGTCGGGTGGTATGGAG.....AA 8004

94 sGlyIleLysLysLeuLeu.....GlyAspV 103
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8005 AGGCTGGAGCAGGTGTCGCCGGTCTCCGGTGTCTCGAGTCAGCGCGGG 8054

103 alGlnMetProValAlaAlaValValMetAlaCysSerArgThrAsp 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8055 CCTAGTGTCCGATCGATGCC.....GACCTACCGCGCGGAGCGTATCCAC 8098

120 TyrLeu.....GluArgThrIleLysSerIleLeuLysTy rGlnThrSe 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8099 TACCTCTCGATGGTGAGTGAAGATCGTGCTGACG....CAGGCATG 8145

134 rValAlaSerLys.....TyrPro.....LeuPheIleS 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8146 GCTGGATGCAAACTGTCATGCGCGCGGGGATCCAGCGGCTCTCTGTGA 8195

144 erGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSerTy rGly 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8196 CGATGCCGGCTCGAAGCGCAGCGACCGACTCCGATGATGCCCAT 8245

161 GlnLeu...ThrTyrMetGlnHisLeuAspTy rGluProValHisThrGl 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8246 CAGACACCTTCGGATCTCGCGTATGCTATCTACACCTCGGGATCCACAG 8295

176 uArgProGlyGluLeuValAlaTy rTyrLysIleAlaArgHisTy rLysT 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8296 GTTGCCCAAGGGGTGATGATCGATCATCGGGGTGCCGTCAACACCATCC 8345

193 rPalAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIle 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8346 TGGACATCAACAGCGCTTCGAATAGGCCCGGAGACAGACTGTGGCG 8395

210 LeuGluAsp.....AspMetGluIleAlaAlaAspPhePheAspTy r 223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8396 CTCCTCCCTCGCTCAGCTTCGATCTCTCGGTC...TAGCATGTGTCGGAT 8442

223 rPheGluAlaGlyAlaThr.....GlyAspV 229
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8443 CCTGGCGCGGGGTGATGATCGATCGGTCGCCGACCGCTCCAAAGTCGCGG 8492

230 .....LeuLeuAspArgAspLysSerIleMetAla 239

```


9359GTTCCGGCGGAGCGG.....GGCTGGAGTCTACGCG 9391

424 PheGlnThrSerArgArgValpHeLeuValSerProAspSerLeuArgGI 440

9392 CQTCCCGCTACGCTCCGACGTTCTCTGAGGCCCGATTCCATTGTTGA 9441

440 nLeuGly 442

9442 ATTCGCG 9448

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-904-452-3

```

seq_documentation_block:
  ; Sequence 3, Application US/08904452
  ; Patent No. 6083742
  ; GENERAL INFORMATION:
  ; APPLICANT: Randazzo, Filippo M.
  ; TITLE OF INVENTION: Mammalian Deep Orange Proteins
  ; NUMBER OF SEQUENCES: 4
  ; CORRESPONDENCE ADDRESS:
  ; ADDRESSEE: Chiron Corporation
  ; STREET: 4560 Horton Street
  ; CITY: Emeryville
  ; STATE: California
  ; COUNTRY: U.S.A.
  ; ZIP: 94608-2916
  ; COMPUTER READABLE FORM:
  ; MEDIUM TYPE: Floppy disk
  ; COMPUTER: IBM PC compatible
  ; OPERATING SYSTEM: PC-DOS/MS-DOS
  ; SOFTWARE: PatentIn Release #1.0, Version #1.30
  ; CURRENT APPLICATION DATA:
  ; APPLICATION NUMBER: US/08/904,452
  ; FILING DATE: 31-JUL-1997
  ; CLASSIFICATION: 435
  ; ATTORNEY/AGENT INFORMATION:
  ; NAME: Potter, Jane E.R.
  ; REGISTRATION NUMBER: 33,332
  ; REFERENCE/DOCKET NUMBER: 1240.004
  ; TELECOMMUNICATION INFORMATION:
  ; TELEPHONE: (510) 923-2718
  ; TELEFAX: (510) 655-3542
  ; INFORMATION FOR SEQ ID NO: 3:
  ; SEQUENCE CHARACTERISTICS:
  ; LENGTH: 4602 base pairs
  ; TYPE: nucleic acid
  ; STRANDEDNESS: single
  ; TOPOLOGY: linear
  ; MOLECULE TYPE: DNA (genomic)
  US-08-904-452-3

```

alignment_scores:		
Quality:	95.50	Length: 311
Ratio:	0.637	Gaps: 16
Percent Similarity:	48.232	Percent Identity: 23.794

alignment_block:

US-09-591-466C-2 x US-08-904-452-3

Align seg 1/1 to: US-08-904-452-3 from: 1 to: 4602

23 TyrIleGlnMetArgLeuPheAlaThrGlnSerGluTyrVal..... 36
||||| :||| :||| :||| :||| :|||
2139 TACTGGACATGAACCGCTTTGACCTGCCAAAGAGATTGTACAGAGCG 2188

37AspargLeuAlaAlaLalleGluAlaGlu.....AsnHisCysThrs 51
||| ||| :||| :||| :||| :||| :|||
2189 GCCTGATTGCCTGCACACGGTCCTGCCGAGAGCGTGATTCTGTTC 2238

51 erGlnThrArgLeuLeuleaspLysIleserGlnGlnGlnGlyArgVal 67
||| ||| :||| :||| :||| :||| :|||
2239 CCCAGCATCGTACCTG.....GAGAGCGCCCGCTGC 2270

58	ValAlaLeuGlu.....	GluGlnMetLysHisGlnAspGlnI	80
2271	TATGGCGCTGACACAGAGACTATTTGAGGAGATGCGCCTCAAGTTCTTGA		2320
80	uCySAArgLeuArgAlaLeuValGlnAspLeuGluSerLysGlyIleL		97
2321	GCCTGGCAAGAGGAGGCGCTGGCGAGTTCTCCAG.....		2358
97	ysLysLeuIleGlyAspValGlnMetProValAlaAlaValValMet		113
2359	GGAACTGGCCGGCTTGAAGCCGACGAGCGGTACCCAGGCCACACTGCTG		2408
114	AlaCysSerArgThrAsp...TyrLeuGluArg.....		123
2409	ACCACCTTGGCTGACAGAGCTCTACCTGAGCCGCTGGTGGCTCTGCAGG		2458
124	ThrIleLysSerIleLeuL	130
2459	TGACCCAGATGCTGTGACTCTCTACCGGACACACGGGAGTGTTCCGTA		2508
130	ysTyrGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAsp		146
2509	CTTTCTCAGTAGCCCTCGGCACAAAGAGTGCTCTT.....		2546
147	GlySerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuTh		163
2547	GCTACGCGGGCTCTATCCACGAGTGCTCGCCAGTCACGAGACACACA		2596
163	rTyrMetClnHisLeu.....	AspTyrGluProVal.....	173
2597	GCACATGGTTATTTTGGAGTGATCATCGAGAGCTATGAACGGGTGGTG		2646
174HisThrGluArgProGlyGluLeuValAlaTyrTyrLysIleAla		188
2647	CATACCACTGTCACCATGAGCTTACGAGGAGGCCCTGGCTGCTGGTGC		2696
189	ArgHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSe		205
2697	CGCCACCGGGAC.....	CCCGAGCTCTCTACAAA.....	2731
205	rArgValIleLeuGluAspMetGluIleAlaAlaAspPhePheA		222
2732	CCCCATTCTCATC.....	CGCCACATCCCCCGCAGCTCGTAG	2769
222	spTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMet		238
2770	ACGCTGGATTGAGATGGGACGGCTGGATGCTCGCACCTCATCCCT		2819
239	AlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspP		255
2820	GCCTTGGTGAACCTACAGCCAGGCGGTGAGGCCACGAGGTGAGCCAGGC		2869
255	roAspAlaLeuTyrArgSerAspPheProGlyLeuGlyTrpMetLeu		271
2870	CATCCGCTACATGGAATCTGCGTGAATGTGCTGGTG.....	AGACGG	2913
272	SerLysSerThrTrpSerGluLeuSerProLysTrpProLysAla.....		286
2914	AGCAGGCCATTCAACACTACCTGCTGCTCCCTGTATGCCCGTGGCCACCA		2963
287TyrTrpAspAspTrpLeuArgLeu		294
2964	GCCTCACTTCTGGCATACCTGGAACCAAGTGT		2994

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-123-912-111

seq_documentation_block:

; Sequence 111, Application US/09123912A

; Patent No. 6312695

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tongtong

```

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
; FILE REFERENCE: 210121.455C1
; CURRENT APPLICATION NUMBER: US/09/123.912A
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: 09/040.802
; PRIOR FILING DATE: 1998-03-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 1419
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-123-912-111

```

```

alignment_scores:
    Quality: 94.00      Length: 382
    Ratio: 0.563        Gaps: 17
    Percent Similarity: 43.717    Percent Identity: 17.801

alignment_block:
US-09-591-466c-2 x US-09-123-912-111 ..

Align seg 1/1 to: US-09-123-912-111 from: 1 to: 1419

64  GlnGlyArgValAlaLeuGluGlnMetLysHisGlnAspGlnG1 80
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
244 CGAGGAGCCACCGCTCCCGATTGGAGGAGGTGTTTCACTCTGAAAAAGA 293

80  ucysArgGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyIleL 97
    | ::::: ::::: ::::: ::::: ::::: ::::: :::::
294 GACGAGAGCTCAGAATAAAGCGTGAAGAAAAGAGGTGGTAAGAAATAA 343

97  ysLysLeuIleGlyAspValGlnMetProValAlaAlaValValMet 113
    || ::::: ::::: ::::: ::::: ::::: :::::
344 AGGCTGAAGGAAAAGAGATTGAG..... 366

114 AlaCysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuL 130
    ::::: ::::: ::::: ::::: ::::: ::::: :::::
367 .....AACAGAGACGATACATCAACAAATCCAAAAGTTTTTGAC 407

130 styGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspG 147
    ::::: ::::: ::::: ::::: ::::: :::::
408 TCAATAAAGCAAACTCACATGATTATGAACCTGAACATAACCAAC.... 453

147 lysSerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThr 163
    ::::: ::::: ::::: ::::: ::::: :::::
454 .....AGGCTGTTTGGAGAAAAAACA 474

164 TyrMet.....GlnHisLeuAspTyrGluProValHisThrGluAr 177
    ::::: ::::: ::::: ::::: ::::: :::::
475 TACCTCTTCTTCAAAATACCTAGATTAT..... 504

177 gProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrLysTrpA 194
    ::::: ::::: ::::: ::::: ::::: :::::
505 .....GTTGAAAAATATTATCATGCAT 526

194 laLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleLeu 210
    ::::: ::::: ::::: ::::: ::::: :::::
527 CTCTGGAACCTGTT..... 540

211 GluAspAspMetGluIleAlaAlaAspPheAspTyrPheGluAlaG1 227
    ::::: ::::: ::::: ::::: ::::: :::::
541 .....GATTTTGTAAATGCAGC 557

227 yAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerSerTrpA 244
    ::::: ::::: ::::: ::::: ::::: :::::
558 C.....GATGAAGTCGAAGAAGATTAAATTCCTGGG 589

244 snAspAsnGlyGlnArgGlnPheValGlnAsp.....ProAspAlaLeu 258
    ::::: ::::: ::::: ::::: ::::: :::::
590 TTGAAAGCAAAACAATGAAAAATCAAGGACTTGTCCCGAGATGGCTCT 639

```

```

259 TyrArgSerAspPheProGlyLeuGlyTyrMetLeuSerLysSerTh 275
    |||
640 ATTAGTAGC..... 648

275 rTrpSerGluLeuSer.ProLysTrpProLysAlaTyrTrpAspAspTrp 291
    ||| |||::: |||
649 .....TCTACCAAGCTGG.....TGCTGGTCAACATGG 676

292 LeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVa 308
    ::::: ||||| ||| |||
677 TTTATTTTAAAGGGCAATGGACAGGGAGT.....TTAAGAAAGAAAT 720

308 lCysArgThrTyrAsnPheGly..GluHisGlySerSerLeuGly.... 322
    : ||| ||||| ::::: ::::: |||
721 ACTAAGGAAGAGAAATTTGGATGAATAAGACACAAAGTAATCTGTACA 770

323 .....GlnPhePheLysGlnTyrLeuGluProIleL 333
    ::::: ::::: ::::: :::::
771 GATGATGACACAGAGCATTCCTTTAGCTTCACTTTCTGGAGGACTTGC 820

333 ysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuSer...Tyr 348
    ::::: ::::: ::::: ::::: ||||| |||
821 AGGCCAAATTTCTAGGGATTCCATATAAAACACACCTAACCATGTTT 870

349 LeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAl 365
    ::::: ||| ::::: :::::
871 GTGCTTCTGCCCAACGACATCGATGGCTGGAGAGATAATAGATAAAAT 920

365 aLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspG 382
    ::::: ||| ::::: :::::
921 AGTCTCT.....GAGAAATTTGGTAGAGTAGTCCAG 955

382 lyAspVal.....ArgIleGlnTyr 388
    || :::::
956 GCATATGGAAGAAAGAAAGTGAATCTGCACCTGCCCGGTTTGAGGTG 1005

389 ArgAspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheG1 405
    ::::: |||::: |||:::
1006 GAGGACAGTTAGCATCTAGAGCGGCTCGCTGCCATGGGATGGGCGA 1055

405 uGluTrpLysAspGlyValProArgAlaAlaTyrLysGlyIle 419
    : ::::: ::::: ||| |||:::
1056 TGCTTCACTGAG.....CACAAAGCCGACTACTCGGAATG 1092

seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-104-068-3

seq_documentation_block:
; Sequence 3, Application US/09104068
; Patent No. 6238882
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Lenox, Anna L.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; TITLE OF INVENTION: g1d1
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,068
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051,379
; FILING DATE: 01-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1871 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-104-068-3

```

```

alignment_scores:
    Quality: 92.50      Length: 462
    Ratio: 0.465        Gaps: 19
    Percent Similarity: 43.074    Percent Identity: 17.100

```

```
alignment_block:

```

```

US-09-591-466c-2 x US-09-104-068-3 ..
Align seg 1/1 to: US-09-104-068-3 from: 1 to: 1871

23 TyrIleGlnMetArgLeuPhe..... 29
|||||.....
196 TATCATCAGATGAAGTCTCAACAGGAGGAGGCCCGCCCGCTTCGTGC 245
30 .....AlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
|||||.....
246 CCTCGTCGCGAGCGTGATGAAGCACTTACTCTAAGAAATCGCAAGA 295
42 laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuIleAsp 58
|||||.....
296 CAGTTCGAAATCAAGAAATCTGACCTTCGT....CAAACCATGATTGAT 342
59 LysIleSerGlnGlnGlnArgValValAlaAlaLeuGluGlnMetIly 75
|||||.....
343 GAGATTTCGTGGAGATGGCAAGTTGTCGGTGTG...CGTACAGCCAC 389
75 sHisGlnAspGlnGluCysArgGlnLeu..... 84
|||||.....
390 CCATCAAGAAATATGCTGCTAAGCGCTGTATTGTGACGACGGGACTGCTC 439
85 .....ArgAlaLeuValGlnAspLeuGlu...SerLysGlyIleLys 97
|||||.....
440 TCGTGGGAAATTTATCATCGGAGACCTCAAGTACTCATCAGTTCTAAC 489
98 LysLeuIleGlyAspValGlnMetProValAlaAlaValValMetAl 114
|||||.....
490 CACAGCTTGGCTTCTATTAACTCA..... 513
114 aCysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLys 131
|||||.....
514 .....GCTGACAATCTCAAGGAACCTGGGTCTCGAATCGGTCTGT 553
131 yrGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGly 147
|||||.....
554 TCAAGACA.....GGA 564
148 SerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTy 164
|||||.....
565 ACCCTCCAGCGTCAAGGCTTCTTCTATCAATTAC...GATGACCGGA 611
164 rMetGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluL 181
|||||.....

```

```

612 AATTACCCAGGACGACGAGTGCCTAATCAT..... 642
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGln 197
642 ..... 642
198 LeuPheHisLysHisAsnPheSerArgValIleIleLeuGluAspAspMe 214
642 ..... 642
214 tGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuL 231
|||||.....ACTTCAC 658
643 .....TTCTCATAC..... 658
231 euAspArgAspLysSerIleMetAlaIleSerSerTrp.....Asn 244
|||||.....
659 GTGATCAGGATTATGTCAAAGATCAAGTGCATGCTCGTTGACCTATACC 708
245 AspAsnGlyGlnArgGlnPheValGlnAspProAspAlaIleTyrArgSe 261
|||||.....
709 AATGGTACCAGTCATGAGATTATCCAA.....ACAACCTCCACCGTGC 752
261 rAspPhePheProGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerG 278
|||||.....
753 GCCTATGTTTACAGGTGTG.....GTCAAGG 778
278 luLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeu 294
|||||.....
779 GAGTGGGCGCTCGTACTCGCTGATTGAAGACAAAGATTGTGCGCTTT 828
295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
|||||.....
829 GCGGACAGGAACGTCACCAACTCTCTCTTGAGCCAGAA..... 867
311 rTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnT 328
|||||.....
868 .....GGACGCAATACTGAGGAAGTCTATGTTCAG 898
328 yrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMet 344
|||||.....
899 GACTTTCAACCAAGTCTGCTGAGGATGTCAG..... 930
345 AspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLe 361
|||||.....
931 .....CGTGACTT 938
361 uValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlap 378
|||||.....
939 GGTTCATTCCATCAAGAGTTTGGAAATGCAGAGATGATGCGGACAGGTT 988
378 heAsnIleAspGlyAspValArgIle.....GlnTyrArgAspGlnLeu 392
|||||.....
989 ATGTATTGAGTATGATATGTTGCTGCTCATCATGTCGTCGCGACTTTG 1038
393 AspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys... 408
|||||.....
1039 GAAACCAAGAAATCTCA.....GGTCTCTCTACTGCTGCTGCAGAC 1079
409 .....AspGlyValProArgAlaAlaTyrLysGlyIleVal.... 420
|||||.....
1080 AAATGGAACATCAGGTTATGAAGAGCTGTGCGCAAGGGATTATCGCGG 1129
421 .....ValPheArgPheGlnThrSerArgArgValPheLeuVal 433
|||||.....
1130 GTATCAATGGCGCTCTGAAATCCAAGTAAACCTGAGTTGATTCTAAAA 1179
434 SerProAspSerLeuArgGlnLeuGlyValGluAsp 445
|||||.....
1180 CGAAGTGCAGGTTATATCGGGGTGATGATCGACGAC 1215

```

```
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-09-104-068-1
```

```
seq_documentation_block:
```

```

; Sequence 1, Application US/09104068
; Patent No. 6238882
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard
; APPLICANT: Palmer, Leslie M.
; APPLICANT: Fedon, Jason C.
; APPLICANT: Lenox, Anna L.
; APPLICANT: Wang, Min
; APPLICANT: Jaworski, Deborah D.
; TITLE OF INVENTION: gldal
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,068
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051,379
; FILING DATE: 01-JUL-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-09-104-068-1

alignment_scores:
    Quality: 92.50      Length: 462
    Ratio: 0.465        Gaps: 19
    Percent Similarity: 43.074    Percent Identity: 17.100

alignment_block:
US-09-591-466c-2 x US-09-104-068-1 ..
Align seg 1/1 to: US-09-104-068-1 from: 1 to: 2100

23 TyrIleGlnMetArgLeuPhe..... 29
334 TATATCCAGATGAAGTCTCAACACAGGAGGCCCGCCGCTTCGTGC 383
30 .....AlaThrGlnSerGluTyrValAspArgLeuAlaAa 42
384 CCTTCGTGGCGGCTGATAGGAAGTTACTCTAAGGAATGCCCAAGA 433
42 laileGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleasp 58
434 CAGTTGAAATCAAGAAATCTGACCCCTCGT...CAAACCATGATTGAT 480
59 LysIleSerGlnGlnGlnGluArgValValAlaLeuGluGlnMetLy 75
481 GAGATTTTGGTGAAGATGGCAAGTTGTCGTGTG...CGTACAGCCAC 527
```

```

75 sHisGlnAspGlnGluCysArgGlnLeu..... 84
|||||: : : : :
528 CCATCAAGAATATCTGCTAAGGCTGTATTGTGACGACGAGGACTGCTC 577
85 .....ArgAlaLeuValGlnAspLeuGlu...SerLysGlyIleLys 97
: : : : : |||||: : : : :
578 TCCGTGGGAAATTAATCATCGGAGACCTCAAGTACTCATCAGGTCTTAAC 627
: : : : : |||||: : : : :
98 LysLeuIleGlyAspValGlnMetProValAlaAlaValValMetAl 114
: : : : : |||||: : : : :
628 CACAGCTTGGCTTCTATTAACTA..... 651
114 aCysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLys 131
: : ||| |||||: : : : :
652 .....GCTGACAAATCTCAAGGAAGTGGGTCTCGAAATCGTGCT 691
131 YrGlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGly 147
: : : : : |||||: : : : :
692 TCAGACA.....GGA 702
148 SerAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTy 164
: : ||| |||||: : : : :
703 ACCCTCCACGCTGCAAGGCTTCTTCTCAATTAC...GATGTGACGGA 749
164 rMetGlnHisLeuAspTyrGluProValHisThrGluArgProGlyGlu 181
: : ||||| |||||: : : : :
750 AATTCAAGCCAGGAGACGAAGTGCCTAATCAT..... 780
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGln 197
780 ..... 780
198 LeuPheHisLysHisAsnPheSerArgValIleLeuLeuGluAspPme 214
780 ..... 780
214 tGluIleAlaAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuL 231
|||||: : : : :
781 .....TTCATAC.....ACTTCA 796
231 euAspArgAspLysSerIleMetAlaIleSerSerTrp.....Asn 244
|||||: : : : :
797 GTGATGAGGATTATGTCAAAGATCAAGTGCATGCTGGTGCACCTATACC 846
245 AspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeuTyrArgSe 261
: : : : : |||||: : : : :
847 AATGTTACCAGTCATGAGATTATCCAA.....AACAACTCCACCGTGC 890
261 rAspPhePheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG 278
: : ||||| |||||: : : : :
891 GCCTATGTTTACAGGTGTG.....GTCAAG 916
278 luLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeu 294
: : |||||: : : : :
917 GAGTGGGGCTCGTTACTGCTCCGTCGATTGAAGACAAGATTGCGCCTTT 966
295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
: : |||||: : : : :
967 GCGGACAAGGAACGTCACCAACTCTCTTGTGAGCCAGAA..... 1005
311 rTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnT 328
|||||: : : : :
1006 .....GGACGCAATTAAGGAAGTCTATGTGTCAAG 1036
328 YrLeuGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMet 344
|||||: : : : :
1037 GACTTTCAACCAAGTGCCTGAGGATGTCAG..... 1068
345 AspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLe 361
: : |||||: : : : :
1069 .....CGTGACTT 1076
361 uValLysLysAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaP 378
```

```

1077 GGTTCATCCATCAAGGTTGGAAATGCGAGATGATCGGACAGGTT 1126
      |||  ... |||  .....:||||:  ...  ....
378 heAsnIleAspGlyAspValArgIle.....GlnTyrArgAspGlnLeu 392
      ::  |||:  |||:  ...  |||  |||  |||
1127 ATGCTATTGAGTATGATATGCTTGGCTTCATCATGCTGCGTGGCGACTTG 1176
      :  |||:  |||:  ...  |||  |||  |||
393 AspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys... 408
      ::  ::  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
1177 GAAACCAAGAAATCTCA.....GCTCTCTCACTGCTGGTGCAGAC 1217
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
409 .....AspGlyValProAlaAlaTyrLysGlyIleVal.... 420
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1218 AATGCAACATCAGGTTATGAAGAAGCTGCTGGCCAAAGGATTATCGCGG 1267
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
421 .....ValPheArgPheGlnThrSerArgArgValPheLeuVal 433
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1268 GTATCAATCGCGCTCTGAAATCCAAAGGTAAACCTGAGTTGATTCTAANA 1317
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
434 SerProAspSerLeuArgGlnLeuGlyValGluAsp 445
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1318 CGAAGTGACGGTTATATCGGGGTGATGATCGAGC 1353
      |||:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-617-785-11

seq_documentation_block:
; Sequence 11, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttnher, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 2745
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2745)
; US-08-617-785-11

alignment_scores:
Quality: 91.50 Length: 270
Ratio: 0.732 Gaps: 14
Percent Similarity: 46.296 Percent Identity: 22.222

alignment_block:
US-09-591-466C-2 x US-08-617-785-11 ..
Align seg 1/1 to: US-08-617-785-11 from: 1 to: 2745
44 GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuAspLysII 60
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1153 GACACAGATCCAAATGCACAGGACAGGAGAGATTGGAAGAAAGATTCCAA 1202
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
60 eserGlnGlnGlnGlyArg.....ValValAlaL 70
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

1203 CTATCAGCAGGAGGGTAAAGTCCAGTTCGTGATTGACGCCAGTCTATGCTA 1252
70 euGluGluGlnMetLysHisGlnAspGlnGlu...CysArgGlnLeuArg 85
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1253 TGGCTCAGCCCTTCACCATATGAACAGGATCTCTGTGCTGACTACCGG 1302
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
86 AlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIleGlyAs 102
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1303 GGTGTCTCCCGAGAGATGGAGCAAGCTGGAGGCAAGAAGTTGCTGAAGTA 1352
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
102 pValGln.....MetProValAlaAlaValValMetAlaC 115
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1353 TATAGCAATGTTTCAATGTTAGTGTGGCACTCCAGTGATGTTTA 1402
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
115 ysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyr 131
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1403 ACAAGAACGGGATGCACCTGGCGT.....TATGACATCTTTCAGTAC 1446
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
132 GlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspLys 148
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1447 CAGACCACA.....AACACCAG 1463
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
148 tAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrM 165
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1464 CAACCCGGGTTACCGTCTGATC.....GGCAGTGGACAGACG 1501
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
165 etGlnHisLeuAspTyrGluProVal.His.....ThrGluAr 177
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1502 AACTTCAGCTCAATATAGAAGACATGCAGTGGGTAAAGGAGTCCGAGAG 1551
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
177 gProGlyGluLeuValAlaTyrTyrLysIleAla.....A 189
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1552 ATACCCGGCTCAGTGTGCACACTACCATGTAAAGCCAGGACAGAGAA 1601
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
189 rGHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSer 205
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1602 GACACAGAAAG.....GAACCTCTGCTGTTGGAGCTGTGAGC 1639
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
206 ArgValIleIleLeuGluAspAspMet.....GluIleAl 217
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1640 CTGTCGATGTTTACCATGACCATGTTTGATGAGATGACATGCCAGCATTC 1689
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1690 CCT.....ATGACCAGAGGGCCCAATGAAATCGAACCGG 1724
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
234 sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1725 ATGCCAGGATATCCCATCATCAAACTGGAGTGGCACTCCC..... 1765
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1766 .....CCTGGGCT 1773
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerPro.....L 282
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1774 GTATTCTCTCTCTGCGCAATGTTGGGATCATTGCCACCATCTTTGT 1823
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
282 yStrpPro 284
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
1824 CATGGCCA 1831
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
seq_name: /cgn2_6/ptodata/2/1na/6B_COMB.seq:US-08-617-785-13

seq_documentation_block:
; Sequence 13, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttnher, Irene
; APPLICANT: Knopfel, Thomas

```

;; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,

;; FILE REFERENCE: 4-19679/A/PCT

;; CURRENT APPLICATION NUMBER: US/08/617,785E

;; EARLIER FILING DATE: 1996-03-19

;; EARLIER APPLICATION NUMBER: PCT/EP94/02991

;; EARLIER FILING DATE: 1994-09-07

;; EARLIER APPLICATION NUMBER: EPO 9416553.7

;; EARLIER FILING DATE: 1994-08-19

;; EARLIER APPLICATION NUMBER: EPO 93810663.0

;; EARLIER FILING DATE: 1993-09-20

;; NUMBER OF SEQ ID NOS: 26

;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 13

;; LENGTH: 2766

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)..(2766)

US-08-617-785-13

alignment_scores:

Quality: 91.50 Length: 270

Ratio: 0.732 Gaps: 14

Percent Similarity: 46.296 Percent Identity: 22.222

alignment_block:

US-09-591-466c-2 x US-08-617-785-13 ..

Align seg 1/1 to: US-08-617-785-13 from: 1 to: 2766

44 GluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAspLysI1 60

1153 GACACAGATCGCAAAATGACAGGACAGAGAGAAATGGAAAGATTCCAA 1202

60 eSerGlnGlnGlyArg.....ValValAlaL 70

1203 CTTATGAGCAGGAGGGTAAAGTCCAGTTCGTGATTGACGCAGTCTATGCTA 1252

70 euGluGlnMetLysHisGlnAspGlnGlu...CysArgGlnLeuArg 85

1253 TGGCTCACGCCCTTCACCATCATGAACAAGGATCTCTGTCTGACTACCGG 1302

86 AlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLysIleGlyAs 102

1303 GGTGCTGCCACAGATGCAGCAAGCTGGAGGCAAGAAGTTGCTGAAGTA 1352

102 pValGln.....MetProValAlaAlaValValMetAlaC 115

1353 TATACCCAATGTAAATTTCAATGGTAGTGTCTGGCAGTCCAGTGATGTTA 1402

115 ysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyr 131

1403 ACAAGAACGGGGTCCACCTGGCGGT.....TATGACATCTTTCAGTAC 1446

132 GlnThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySe 148

1447 CAGACCACA.....AACACCAG 1463

148 rAsnProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrM 165

1464 CAACCCGGGTACCGTCTGATC.....GGCAGTGGCAGCAGCAGC 1501

165 etGlnHisLeuAspTyrGluProVal.His.....ThrGluAr 177

1502 AACTTCAGCTCAATATAGAAGACATGCGTGGGGTAAAGGAGTCCGAGAG 1551

177 gProGlyGluLeuValAlaTyrTyrLysIleAla.....A 189

1552 ATACCGCGCTCAGTGTGCACATACCATGTAAGCCAGGACAGAGAAAGAA 1601

189 rGHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSer 205

1602 GACACAGAAAG.....GAACCTCCTTCTGCTTTGGACCTGTGAGC 1639

206 ArgValIleIleLeuGluAspAspMet.....GluIleAl 217

1640 CTTGCGATGGTTACCAGTACCAGTTTGTGATGAGATGACATGCCAGCATTGC 1689

217 aAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234

1690 CCCT.....ATGACCAGAGGCCCAATGAAAAATCGAAACCGG 1724

234 sPlysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250

1725 ATGCCAGGATATTCATCATCAAACTGGAGTGGCACTCCC..... 1765

251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267

1766CCTGGGCT 1773

267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerPro.....L 282

1774 GTGATTCTCTCTCTCGCAATGTGGGGATCATTTGCCACCATCTTTGT 1823

282 ysTrpPro 284

1824 CATGGCCA 1831

OM of: US-09-591-466c-2 to: N_Geneseq_032802:* out_format : pfs
Date: Aug 13, 2002 10:47 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame1.p2n.model -DEV=xlp
-O=cpn2_1/USPTO.spool/US09591466/runat_12082002_095313.4300/app_query.fasta_1.511
-DB=N_Geneseq_032802 -QFMP=fastap -SUFFIX=ring -GAPOP=12_000
-GAPEXT=4_000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4_500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6_000 -FGAPEXT=7_000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6_000 -DELEXT=7_000 -START=1 -MATRIX=bycosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09591466 -CCNL_1_625 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-591-466c-2
Query length: 446
Database: N_Geneseq_032802.*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 268.160000

WARN: APPL_compute_dev_space_s: pointers violation: prPtr=5ble76c !

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID55/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AA78001 +		2345.00	4502.95	1.8e-242	16
/SID55/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AA78002 +		2198.00	4218.34	1.3e-226	17
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA39340 +		1776.00	3403.20	3.3e-181	16
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA39340 +		1757.50	3367.42	3.2e-179	16
/SID55/gcgdata/geneseq/geneseq-embl/NA1994.DAT:AA062625 +		735.50	1386.83	6.7e-69	25
/SID55/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AA78003 +		727.00	1388.31	5.6e-69	51
/SID55/gcgdata/geneseq/geneseq-embl/NA1992.DAT:AA25594 +		726.50	1369.75	6.0e-68	24
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA19837 +		723.50	1370.85	5.2e-68	32
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA25595 +		719.50	1353.30	5.0e-67	32
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB22693 +		669.50	1086.68	2.3e-61	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA78002 +		579.50	1086.68	3.5e-52	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB22692 +		506.00	936.36	8.3e-44	4
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA69652 +		469.00	889.03	3.6e-41	53
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA15975 +		363.00	669.04	6.4e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA160103 +		363.00	666.08	9.4e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA37048 +		363.00	666.03	9.5e-29	27
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA321483 +		363.00	666.03	9.5e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA202922 +		363.00	666.03	9.5e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA78002 +		363.00	666.03	9.5e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA91482 +		363.00	666.03	9.5e-29	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA294210 +		363.00	666.00	9.5e-29	27
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA158317 +		361.50	664.76	1.1e-28	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA18300 +		359.00	659.84	2.1e-28	23
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA01130 +		246.50	465.33	1.4e-17	30
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA78002 +		215.50	395.60	1.1e-13	7
/SID55/gcgdata/geneseq/geneseq-embl/NA1998.DAT:AA78002 +		114.50	187.79	0.0411	22
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AA58411 +		111.00	179.85	0.1138	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AA58553 +		111.00	179.85	0.1138	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB20745 +		110.50	185.30	0.0566	1
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA24480 +		109.50	181.84	0.0882	16
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA78002 +		109.50	181.84	0.0882	1
/SID55/gcgdata/geneseq/geneseq-embl/NA1997.DAT:AA78002 +		108.00	179.97	0.1122	14
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA20584 +		107.00	185.13	0.0579	7
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA20584 +		106.50	158.16	1.84	6
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA14503 +		105.00	158.12	1.85	64
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA18405 +		105.00	158.12	1.85	6
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AA59308 +		104.00	165.87	0.6843	2
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB22099 +		104.00	145.74	9.04	1
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB22098 +		104.00	143.59	11.91	1

/SID55/gcgdata/geneseq/geneseq-embl/NA1996.DAT:AA78002 + 102.50 159.77 1.50
/SID55/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AA78002 + 101.00 162.22 1.09
/SID55/gcgdata/geneseq/geneseq-embl/NA1996.DAT:AA78002 + 100.50 161.16 1.25
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AA78002 + 100.00 156.18 2.37
/SID55/gcgdata/geneseq/geneseq-embl/NA2001B.DAT:AB22680 + 100.00 151.46 4.34
/SID55/gcgdata/geneseq/geneseq-embl/NA2000.DAT:AA78002 + 98.50 162.65 1.03

seq_name: /SID55/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AA78001

seq_documentation_block:

ID AAX78001 standard; cDNA; 1669 BP.

XX AAX78001;

XX 19-AUG-1999 (first entry)

XX Potato GntI cDNA.

XX GntI: beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;
XX deficient; defective; detection; transgenic plant; sugar residue;
XX medicine; antigenicity; deglycosylation; potato; ds.

XX Solanum tuberosum.

XX Key Location/Qualifiers

XX CDS 53..1394

XX /*tag= a

XX /product= "GntI"

XX DE19754622-A1.

XX 10-JUN-1999.

XX 09-DEC-1997; 97DE-1054622.

XX 09-DEC-1997; 97DE-1054622.

XX (VSC/)/ VON SCHAEWEN A.

XX Von Schaewen A;

XX WPI; 1999-338905/29.

XX P-PSDB; AAY08888.

XX Nucleic acid encoding plant N-acetylglucosaminyl transferase I

XX useful for generating plants producing glycosylated proteins of low

XX antigenicity

XX Claim 6; Page 16-18; 37pp; German.

XX This invention describes novel plant N-acetylglucosaminyl transferase I
XX (GntI) proteins and their encoding nucleic acids. The nucleic acid
XX sequences of the invention may be used for recombinant production of the
XX encoded proteins, which are then used to raise antibodies (Ab) for
XX identifying plants with deficient or defective production of the protein.
XX They may also be used to detect such plants by hybridization and to
XX isolate related sequences from other plants or to generate antisense or
XX sense constructs for reducing/deleting GntI protein activity in plants.
XX These transgenic plants may be used to produce glycoproteins with
XX minimal, uniform and defined sugar residues. Such glycoproteins are
XX useful in medicine and research, e.g. human glucocerebrosidase for
XX treating Gaucher's disease. Plants which are defective or deficient in
XX production of the GntI protein can be made to produce glycoproteins with
XX minimal, uniform and defined sugar residues, of low antigenicity. Use of
XX these plants eliminates the need for the difficult isolation and
XX deglycosylation of native proteins or preparation in defective animal
XX cells. This sequence represents potato (Solanum tuberosum) GntI encoding
XX cDNA.

XX Sequence 1669 BP; 489 A; 312 C; 387 G; 481 T; 0 other;

alignment_scores:

Quality: 2345.00 Length: 446
Ratio: 5.258 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-591-466C-2 x AAX78001 ..

Align seg 1/1 to: AAX78001 from: 1 to: 1669

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17
53 ATGAGAGGGAACAAGTTTGGCTTGTATTACGGTACCTTCTCGTCGTGGC 102
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
103 TGTCTCGCCCTTCATCATACATACAGATGCGGCTTTCGCGACACAGTCAG 152
34 lutyValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
153 AATATGTAGACCGCTTGTCTGTGCAATTTGAAGCAGAAAAATCATTGTACA 202
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67
203 AGTCAGACCAAGATTGCTTATTGACAAAGATTAGCCAGCAGCAAGGAAGAT 252
67 lValAlaLeuGluGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
253 AGTAGCTCTTGAAGACAAATGAAGCATCAGGACCAAGGAGTGCAGCAAT 302
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
303 TAAGGGCTCTTGTTCAGGATCTTTGAAGTAGGAGGATGAAAGGTTAATC 352
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
353 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTGTTATGCGTTGCAGTCG 402
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
403 TACTGTACTGTGAGAGGACTATTAAATCCATCTTAAATATACCAACAT 452
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
453 CTGTTGCATCAAAATATCTCTTTTCATATCCAGGATGCATCAATCCT 502
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
503 GATGTAAGAAAGCTTGCTTTGAGCTATGTCAGCTGACGCTATATGACAGA 552
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
553 CTTGGATTATGAACCTGTGCATCTACTGAAAGCAGGGAACCTGGTTGCAT 602
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
603 ACTACAAGATGTCACGTCATTACAAGTGGGCAITGGATCAGCTGTTTCAC 652
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
653 AAGCATAAATTTTACCGGTGTTTCACTACTAGAAGATGATATGAAATTCG 702
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
703 TGCTGATTTTTTTGACTATTTTGGGCTGGAGCTACTCTCTTCTTGACAGAG 752
234 spLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
753 ACAAGTCGATTATGGCTATTCTTCTTGGAAATGACAATGGACAAAGGCAG 802
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
803 TTCGTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTTCTCGTCT 852

267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpP 284
853 TGGATGGATGCTTTCAAATCAAACTTGTCCGAACATATCTCCAAAGTGC 902
284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
903 CAAGGCTTACTGGGATGACTGGCTAAGGCTGAAAGAAAAATCACAGAGT 952
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317
953 CGACAAATTTATCGCCAGAGAGTTTGCAGAACGTACAAATTTTGGTGAGCA 1002
317 scLysSerLeuGlyGlnPhePheLysGlnTyrLeuGluProLleLysL 334
1003 TGGTTCTAGTTTGGGGCAGTTTTTTAAGCAGTATCTTGAGCCAATTAAGC 1052
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
1053 TAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1102
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
1103 GAGGACAACTATGTGAACACACTTTGGCGACTTGGTTAAAAAGGCTAAGCC 1152
367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
1153 CATCCACGAGCTGATGCTGTTTGAAGCATTTAACATAGATGTTGATG 1202
384 aIArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
1203 TCGGTATTTCAGTACAGAGACCAACTAGACTTTTGAAGATATCGCTCGACAG 1252
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrIly 417
1253 TTTGGCATTTTGAAGAAATGGAAGGATGGTGACCACGGGACGACATATAA 1302
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuValS 434
1303 AGGGATAGTAGTTTTCCGGTTTCAACATCTAGACGCTGTGTTCTTGT 1352
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
1353 CCCCTGATTCTCTCGACAACTTGGAGTTGAAGATACT 1390

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX78002

seq_documentation_block:

ID AAX78002 standard; cDNA; 1737 BP.

XX AAX78002;

XX AC AAX78002;

XX DT 19-AUG-1999 (first entry)

XX XX Tobacco GntI cDNA.

XX DE GntI; beta-1,2-N-acetylglucosaminyltransferase; glycoprotein; plant;

XX KW deficient; defective; detection; transgenic plant; sugar residue;

XX KW medicine; antigenicity; deglycosylation; tobacco; ds.

XX OS Nicotiana tabacum.

XX XX Key Location/Qualifiers

XX FH 127..1467

XX FT /*tag= a

XX FT /product= "GntI"

XX XX DE19754622-A1.

XX XX 10-JUN-1999.

XX XX 09-DEC-1997; 97DE-1054622.

XX XX 09-DEC-1997; 97DE-1054622.

PR

XX (VSCH/) VON SCHAEWEN A.
 PA Von Schaewen A;
 PI
 PL
 DR WPI: 1999-338905/29.
 DR P-PSDB; AAY08889.
 XX
 PT Nucleic acid encoding plant N-acetylglucosaminyl transferase I
 PT useful for generating plants producing glycosylated proteins of low
 PT antigenicity
 XX
 PS Claim 7; Page 22-24; 37pp; German.
 XX
 CC This invention describes novel plant N-acetylglucosaminyl transferase I
 CC (GntI) proteins and their encoding nucleic acids. The nucleic acid
 CC sequences of the invention may be used for recombinant production of the
 CC encoded proteins, which are then used to raise antibodies (Ab) for
 CC identifying plants with deficient or defective production of the protein.
 CC They may also be used to detect such plants by hybridization and to
 CC isolate related sequences from other plants or to generate antisense or
 CC sense constructs for reducing/deleting GntI protein activity in plants.
 CC These transgenic plants may be used to produce glycoproteins with
 CC minimal, uniform and defined sugar residues. Such glycoproteins are
 CC useful in medicine and research, e.g. human glucocerebrosidase for
 CC treating Gaucher's disease. Plants which are defective or deficient in
 CC production of the GntI protein can be made to produce glycoproteins with
 CC minimal, uniform and defined sugar residues, of low antigenicity. Use of
 CC these plants eliminates the need for the difficult isolation and
 CC deglycosylation of native proteins or preparation in defective animal
 CC cells. This sequence represents tobacco (Nicotiana tabacum) GntI encoding
 CC cDNA.
 XX
 SQ Sequence 1737 BP; 510 A; 337 C; 390 G; 500 T; 0 other;

alignment_scores:
 Quality: 2198.00 Length: 446
 Ratio: 5.030 Gaps: 0
 Percent Similarity: 97.982 Percent Identity: 92.825

alignment_block:
 US-09-591-466c-2 x AAX78002 ..

Align seg 1/1 to: AAX78002 from: 1 to: 1737

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValAl 17
 127 ATGAGAGGGGAACAAAGTTTGGCTGTGATTTCGGGTACTCTCTCATCTGGC 176
 17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerg 34
 177 TGCTGTGCGCTTCATCTACACACAGATGCGGCTTTTTCGCACACAGTCAG 226
 34 luTyrValAspArgLeuAlaAlaAlaIleGluAlaGluAsnHisCysThr 50
 227 AATATGCAGATCGCCTTGTGCTGCAATTGAACAGAAATCAATTGTACA 276
 51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
 277 AGCCAGACCAGATTGCTTATTGACCAGATTAGCCTGCAGCAAGGAAGAT 326
 67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
 327 AGTTGCTCTTTGAAGACAAATGAAGCGTCAAGGACCAGGAGTGCAGCAAT 376
 84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
 377 TAAGGGCTCTTGTTCAGCATCTTTGAAGTGAAGGCGCATAAAAAAGTTGATC 426
 101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
 427 GGAATGTACAGATGCCAGTGGCTGTGTAGTTGTTATGCTTGAATCG 476

117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
 477 GGCTGATTACCTGGAAAAGACATTAAATCCATCTTAAATACCAAAATAT 526
 134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
 527 CTGTTGGCGTCAAAATATCTCTTTTCATATCCAGGATGGATCACAATCCT 576
 151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
 577 GATGTCAGGAAGCTTGCTTTGAGCTATGATCAGTCAGCTATATGACGCA 626
 167 steuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
 627 CTGGAATTTGAACCTGTGCATCTGAAAGACAGGGAGCTGATTGCAT 676
 184 yrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
 677 ACTACAAAATTCACGTCATTACAAGTGGCATTTGGATCAGCTGTTTAC 726
 201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217
 727 AAGCATAAATTTAGCGCTGTATCATACTAGAGATGATATGAAATGCG 776
 217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
 777 CCTGATTTTGTGACTTTTGGAGCTGGAGCTACTCTTCTTGACAGAG 826
 234 sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
 827 ACAAGTCGATTTATGGCTATTCTTCTTGAAGTGAATGGACAAATGCGAG 876
 251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
 877 TTGTCCAAAGATCCTTATGCTCTTTACCGCTCAGATTTTTTCCCGTCT 926
 267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpP 284
 927 TGGATGGATGCTTTCAAAATCTACTTTGGGACGAAATTTATCTCCAAAGTGC 976
 284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
 977 CAAGGCTTACTGGGACGACTGGCTAAGACTCAAGAGAAATCACAGAGT 1026
 301 ArgGlnPheIleArgProGluValValCysArgThrTyrAsnPheGlyGluHi 317
 1027 CGACAATTTATTCGCCAGAAAGTTTCAGAACATATAATTTTGGTCAGCA 1076
 317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
 1077 TGCTTCTAGTTTGGGCGAGTTTTCAGCAGTATCTTGAGCCCAATTAAC 1126
 334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
 1127 TAATGATGTCAGGTTGATTGGAAGTCAATGGACCTTAGTTACCTTTTG 1176
 351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
 1177 GAGGACAATTTACGTGAACACACTTTTGTGACTTGGTTAAAAAGGCTAAGCC 1226
 367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
 1227 CATCCATGGAGCTGATGCTGCTTGAAGCATTTAATAGATAGATGGTGATG 1276
 384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
 1277 TCGGATTTCAGTACAGATCAACTAGACTTTGAAATATCGCACGGCAA 1326
 401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
 1327 TTTGGCATTTTGAAGATGAAGAGTGGTGTACCACGTCGACGATATAA 1376

417 sGlylleValValPheArgPheClnThrSerArgArgValPheLeuValS 434
|||||
1377 AGCAATAGTAGTTTCGGTACCAACGTCACACGTGATTCTTGTTG 1426

434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
:::|||||
1427 GCCATGATTCGCTCAACACTCGGAATTGAAGATACT 1464

seq_name: /Sids5/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.AAC39540
seq_documentation_block:
ID AAC39540 standard; DNA; 1639 BP.
XX AAC39540;
AC AAC39540;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25006.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0148171.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

```
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 1776.00      Length: 438
  Ratio: 4.429          Gaps: 0
  Percent Similarity: 91.553  Percent Identity: 75.342

alignment_block:
US-09-591-466c-2 x AAC39540 ..
```

```
Align seg 1/1 to: AAC39540 from: 1 to: 1639

9 AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleG1 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 GACTTGAGATTTCTCTCAGCGGAGCTTTCATGTTTCATCTACATCCA 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 GATGAGGCTTTCCAGACGCAATCACAGTATGCAGATCGCCTCAGTTCG 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
42 IaIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 CTATCGAATCTGAGAACCATTTGCACCTAGTCAAAATGCGAGCGCTCATAGAT 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 LysIleSerGlnGlnGlyArgValValAlaLeuGluGlnMetLys 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
200 GAAGTTAGCATCAACAGTCGCGGATTGTGCCCTCGAAGATATGAAGA 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
75 sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 CCGCCAGGACGAAGAACTTGTGCAGCTTAAGGATCTAATCCAGACGTTG 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 LuSerLysGlyIleLysLysLysLeuIleGlyAspValGlnMetProValAla 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 AAAAAAAGGAATAGCAAACTCACTCAAGGTGGACAGATGCCTGTGGCT 349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
109 AlaValValValMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
350 GCTGTAGTGGTTATGGCTGCAGTCGTGCAGACTATCTTGAAGGACTGT 399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 eLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeuP 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 TAAATCAGTTTTAACATATCAAACTCCGTTGCTTCAAAATATCCTCTAT 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 heIleSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSer 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 TTATATCTCAGGATGGATCTGATCAAGCTGTCAAGAGCAAGTCAATTGAGC 499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHisTh 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
500 TATAATCAATTAACACATATGCAGACTTGGATTTTGAACCACTGGTTCAC 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrL 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
550 TGAAGGCGCTGGTGAAGTGCCTGCTACTACAAGATTGCACGCTCACTACA 599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 yStrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIle 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 AGTGGGCACTGGACCAAGTTGTTTTACAAACACAAATTTACTCGAGTGATT 649
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 IleLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheG1 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
650 ATACTAGAAGATGATATGGAATTTGCTCCAGACTTCTTTGATTACTTTGA 699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerS 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 GGCTGCAGCTAGTCTCATGGATAGGATAAAACCAATTATGGCTGCTTCAT 749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 eTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeu 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
750 CATGGAATGATATGACAGACAGCAGTTTGTGCATGATCCCTATGCGCTA 799
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 TyrArgSerAspPhePheProGlyLeuGlyTrpMetLeuSerLysSerTh 275
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
800 TACCGATCAGATTTTTTCTTCCCTGGCTGGTGGATGCTCAAGAGATCGAC 849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
275 rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
850 TTGGGATGAGTTATACCAAAAGTGGCCAAAGGCTTACTGGGATGATGGC 899
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
292 euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

900	TGAGACTAAAGGAAAAACCATAAAGGCCGCCCAATTTCATTCGACCCGGAAGTC	949	PR	23-APR-1999;	99US-0130510.
309	CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePh	325	PR	23-APR-1999;	99US-0130891.
950	TGTAGAACATACAAATTTTGGTGAACATGGTCTAGTTGGGACAGTTTTT	999	PR	28-APR-1999;	99US-0131449.
325	eLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL	342	PR	30-APR-1999;	99US-0132048.
1000	CAGTCAGTATCTGGAACTATAAAGCTAAACGATGTGACGGTTGACTGGA	1049	PR	30-APR-1999;	99US-0132407.
342	ysSerMetAspLeuSerTyrLeuGluAspAsnTyrValLysHisPhe	358	PR	04-MAY-1999;	99US-0132484.
1050	AAGCAAGACCTGGGATACCTGACAGAGGGAACCTATACCAAGTACTTT	1099	PR	05-MAY-1999;	99US-0132485.
359	GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLe	375	PR	06-MAY-1999;	99US-0132486.
1100	TCTGGCTTAGTGAGACAGCAGCAGCAATTCAAGGTTCTGACCTTGCTCT	1149	PR	06-MAY-1999;	99US-0132487.
375	uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL	392	PR	07-MAY-1999;	99US-0132863.
1150	AAAGGCTCAAAACATATAAGGATGATGTTGCTATCCGGTATAAGACCAAG	1199	PR	11-MAY-1999;	99US-0134256.
392	euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys	408	PR	14-MAY-1999;	99US-0134218.
1200	TAGAGTTTGAACGCATTGCAGGGGAATTGGTATATTTGAAGAATGGAAG	1249	PR	14-MAY-1999;	99US-0134221.
409	AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheGl	425	PR	14-MAY-1999;	99US-0134370.
1250	GATGTGTGCTCGAACACGATATAAAGAGGATAGTGGTTCGAATCCA	1299	PR	18-MAY-1999;	99US-0134376.
425	nThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnLeuG	442	PR	19-MAY-1999;	99US-0134941.
1300	GACAAACAGAGCTGTATTCCTGGTTGGGCCAGATTCTCTATATGACGCTG	1349	PR	20-MAY-1999;	99US-0135124.
442	lyValGluAspThr	446	PR	21-MAY-1999;	99US-0135353.
1350	GAATTCGAATTC	1363	PR	25-MAY-1999;	99US-0135629.
seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC50197					
seq_documentation_block:					
ID	AAC50197	standard; DNA; 1641 BP.	PR	28-MAY-1999;	99US-0136782.
XX	AAC50197;		PR	01-JUN-1999;	99US-0137222.
AC			PR	03-JUN-1999;	99US-0137528.
XX	18-OCT-2000	(first entry)	PR	04-JUN-1999;	99US-0137502.
DE	Arabidopsis thaliana	DNA fragment SEQ ID NO: 63936.	PR	07-JUN-1999;	99US-0137724.
KW	Hybridisation assay; genetic mapping; gene expression control;		PR	08-JUN-1999;	99US-0138094.
KW	protein identification; signal transduction pathway;		PR	10-JUN-1999;	99US-0138540.
KW	metabolic pathway; promoter; termination sequence; ss.		PR	10-JUN-1999;	99US-0138847.
XX	Arabidopsis thaliana.		PR	14-JUN-1999;	99US-0139119.
OS			PR	16-JUN-1999;	99US-0139452.
PN	Ep1033405-A2.		PR	17-JUN-1999;	99US-0139453.
XX	06-SEP-2000.		PR	18-JUN-1999;	99US-0139454.
PF	25-FEB-2000;	2000EP-0301439.	PR	18-JUN-1999;	99US-0139455.
XX	25-FEB-1999;	99US-0121825.	PR	18-JUN-1999;	99US-0139456.
PR	05-MAR-1999;	99US-0123180.	PR	18-JUN-1999;	99US-0139457.
PR	09-MAR-1999;	99US-0123548.	PR	18-JUN-1999;	99US-0139458.
PR	23-MAR-1999;	99US-0125788.	PR	18-JUN-1999;	99US-0139459.
PR	25-MAR-1999;	99US-0126264.	PR	18-JUN-1999;	99US-0139460.
PR	29-MAR-1999;	99US-0126785.	PR	18-JUN-1999;	99US-0139461.
PR	01-APR-1999;	99US-0127462.	PR	18-JUN-1999;	99US-0139462.
PR	06-APR-1999;	99US-0128234.	PR	18-JUN-1999;	99US-0139463.
PR	08-APR-1999;	99US-0128714.	PR	21-JUN-1999;	99US-0139817.
PR	16-APR-1999;	99US-0129845.	PR	22-JUN-1999;	99US-0139899.
PR	19-APR-1999;	99US-0130077.	PR	23-JUN-1999;	99US-0140353.
PR	21-APR-1999;	99US-0130449.	PR	23-JUN-1999;	99US-0140354.
			PR	24-JUN-1999;	99US-0140695.
			PR	28-JUN-1999;	99US-0140823.
			PR	29-JUN-1999;	99US-0140991.
			PR	30-JUN-1999;	99US-0141287.
			PR	01-JUL-1999;	99US-0141842.
			PR	01-JUL-1999;	99US-0142154.
			PR	02-JUL-1999;	99US-0142055.
			PR	06-JUL-1999;	99US-0142390.
			PR	08-JUL-1999;	99US-0142803.
			PR	09-JUL-1999;	99US-0142920.
			PR	12-JUL-1999;	99US-0142977.
			PR	13-JUL-1999;	99US-0143542.
			PR	14-JUL-1999;	99US-0143624.
			PR	15-JUL-1999;	99US-0144005.
			PR	16-JUL-1999;	99US-0144085.
			PR	16-JUL-1999;	99US-0144086.
			PR	19-JUL-1999;	99US-0144325.
			PR	19-JUL-1999;	99US-0144331.
			PR	19-JUL-1999;	99US-0144332.
			PR	19-JUL-1999;	99US-0144333.
			PR	19-JUL-1999;	99US-0144334.
			PR	20-JUL-1999;	99US-0144335.
			PR	20-JUL-1999;	99US-0144352.

PR	20-JUL-1999;	9905-01446339;
PR	20-JUL-1999;	9905-01446484;
PR	21-JUL-1999;	9905-01448141;
PR	21-JUL-1999;	9905-01450866;
PR	21-JUL-1999;	9905-01450888;
PR	22-JUL-1999;	9905-01450885;
PR	22-JUL-1999;	9905-01450897;
PR	22-JUL-1999;	9905-01450899;
PR	22-JUL-1999;	9905-01451192;
PR	23-JUL-1999;	9905-01451545;
PR	23-JUL-1999;	9905-01452218;
PR	26-JUL-1999;	9905-01452224;
PR	27-JUL-1999;	9905-01452676;
PR	27-JUL-1999;	9905-01455913;
PR	27-JUL-1999;	9905-01459148;
PR	28-JUL-1999;	9905-01459511;
PR	02-AUG-1999;	9905-01459591;
PR	02-AUG-1999;	9905-01463386;
PR	02-AUG-1999;	9905-01463889;
PR	03-AUG-1999;	9905-01463899;
PR	04-AUG-1999;	9905-01470338;
PR	04-AUG-1999;	9905-01472002;
PR	05-AUG-1999;	9905-01471302;
PR	05-AUG-1999;	9905-01471932;
PR	06-AUG-1999;	9905-01472650;
PR	06-AUG-1999;	9905-01473003;
PR	09-AUG-1999;	9905-01474116;
PR	09-AUG-1999;	9905-01474935;
PR	10-AUG-1999;	9905-01479335;
PR	11-AUG-1999;	9905-01481711;
PR	12-AUG-1999;	9905-01483341;
PR	13-AUG-1999;	9905-01483419;
PR	13-AUG-1999;	9905-01485655;
PR	17-AUG-1999;	9905-01486884;
PR	17-AUG-1999;	9905-01493668;
PR	18-AUG-1999;	9905-01491755;
PR	20-AUG-1999;	9905-01494266;
PR	20-AUG-1999;	9905-01497222;
PR	20-AUG-1999;	9905-01497233;
PR	23-AUG-1999;	9905-01495929;
PR	23-AUG-1999;	9905-01499022;
PR	25-AUG-1999;	9905-01499300;
PR	26-AUG-1999;	9905-01500566;
PR	26-AUG-1999;	9905-01500884;
PR	27-AUG-1999;	9905-01510655;
PR	27-AUG-1999;	9905-01510666;
PR	30-AUG-1999;	9905-01510800;
PR	31-AUG-1999;	9905-01513033;
PR	1-SEP-1999;	9905-01514338;
PR	7-SEP-1999;	9905-01519330;
PR	10-SEP-1999;	9905-01523633;
PR	13-SEP-1999;	9905-01530700;
PR	15-SEP-1999;	9905-01537558;
PR	16-SEP-1999;	9905-01540138;
PR	20-SEP-1999;	9905-01540339;
PR	22-SEP-1999;	9905-01547779;
PR	23-SEP-1999;	9905-01551539;
PR	24-SEP-1999;	9905-01554866;
PR	28-SEP-1999;	9905-01556659;
PR	29-SEP-1999;	9905-01564358;
PR	04-OCT-1999;	9905-01565936;
PR	05-OCT-1999;	9905-01571117;
PR	06-OCT-1999;	9905-01577553;
PR	07-OCT-1999;	9905-01578655;
PR	08-OCT-1999;	9905-01580229;
PR	12-OCT-1999;	9905-01582332;
PR	13-OCT-1999;	9905-01583659;
PR	13-OCT-1999;	9905-01592933;
PR	13-OCT-1999;	9905-01592994;
PR	14-OCT-1999;	9905-01595295;
PR	14-OCT-1999;	9905-01595329;
PR	14-OCT-1999;	9905-01595930;
PR	14-OCT-1999;	9905-01595931;

```

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-016192.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 1757.50
    Ratio: 4.383
    Percent Similarity: 91.136    Percent Identity: 75.227

alignment_block:
US-09-591-466C-2 x AAC50197 ..

Align seg 1/1 to: AAC50197 from: 1 to: 1641

9 AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleG1 25
48 GACTTGAGATTCTCTCATCCCGCAGCTTTTCATGTTTCATCTACATCCA 97

25 nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
98 GATGAGCTTTTTCCAGACGCAATCACAGTATGCAGATGCCTCATCGTCG 147

42 lalleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
148 CTATCGAATCTGAGAACCATTTGCACCTAGTCAATGCGAGCCTCATGAT 197

59 LysIleSerGlnGlnGlyArgValValAlaLeuGluGluGlnMetIly 75
198 GAAGTTAGCATCAACACAGTCGGGATGTTGTCCTCGAAGATATGAAGAA 247

75 sHisGlnAspGlnGlyCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
248 CGCCAGGACGAAGAACTTGTGCAGCTTAAGGATCTAATCCAGACGTTTG 297

92 luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAla 108
298 AAAAAAAGGAATAGCAAAACCTACTCAAGGTGGACAGATGCTGTGCT 347

109 AlaValValMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
348 GCTGTAGTGGTTATGGCTGCAGTCGTCGACGACTATCTTTGAAGGACTGT 397

125 elysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeup 142
398 TAAATCAGTTTAAACATATCAACATCCCGTGTGCTCAAAATATCTCTCTAT 447

142 heIleSerGlnAspLysSerAsnProAspValArgLysLeuAlaSer 158
448 TTATATCTCAGGATGGATCTCATCAAGCTGTCAAGACGAACTCATTCAGC 497

159 TyrGlyGlnLeuThrTyrMet...GlnHisLeuAspTyrGluProValHi 174
498 TATTAATCAATTAACTATATATGCACGACGACTTGATTTTGAACGACGTGT 547

```

Wed Aug 14 11:48:26 2002

```
174 sThrGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisT 191
|||||
548 CACTGAAAGGCGCTGGCGAACTGACGCGTACTACAAAGATTGCACGCTCACT 597
|||||
191 YrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgVal 207
|||||
598 ACAAGTGGGCACTGGACCAAGTGTGTTTACAAACACAAATTTAGTCGAGTG 647
|||||
208 IleIleLeuGluAspAspMetGluIleAlaAlaAspPhePheAspTyrPh 224
|||||
648 ATTATACAGAAATGATATGGAATTTGCTCCAGACTTCTTTGATTACTT 697
|||||
224 eGluAlaGlyAlaThrLeuLeuLeuAspArgAspLysSerIleMetAlaIleS 241
|||||
698 TGAGCTGCACCTAGTCTCATGTGATAGGATAAAACCATATGCTGCTT 747
|||||
241 erSerTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAla 257
|||||
748 CATCATGGAATGATAATGGACAGAAGCAGTTTGTGCATGATCCCTATGCG 797
|||||
258 LeuTyrArgSerAspPhePheProGlyLeuGlyTyrTrpMetLeuSerLysSe 274
|||||
798 CTATACCGATCAGATTTTTTTCCTGGCCTTGGGTGGATGCTCAAGAGATC 847
|||||
274 rThrTrpSerGluLeuSerProLysTrpProLysAla.TyrTrpAspAsp 290
|||||
848 GACTTTGGGATGAGTTATCACCAGAGTGGCCAAAGGCCCTTACTGGGATGAT 897
|||||
291 TrpLeuArgLeuLysGluAsnHisArgGlyArgGlnGlnPheIleArgProG 307
|||||
898 TGGCTGACACTAAAGAAACCATAAAGCGCCCAATTTCATTCGACCGGA 947
|||||
307 uValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnP 324
|||||
948 AGTCTGTAGAACATACATTTTGGTGACATGGGTCTAGTTGGACAGT 997
|||||
324 hePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAsp 340
|||||
998 TTTTCAGTCAGTATCTGGAACCATATAAGCTAAACGATGTGACGGTTGAC 1047
|||||
341 TrpLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHi 357
|||||
1048 TGGAAAGCAAGGACCTGGGTACCTGACAGAGGGAACATAACCAAGTA 1097
|||||
357 sPheGlyAspLeuValLysAlaLysProIleHisGlyAlaAspAlav 374
|||||
1098 CTTTTCCTGCTTAGTGAGACAAGCAGCACCAATTCAAGGTTCTGACCTTG 1147
|||||
374 alLeuLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAsp 390
|||||
1148 TCTTAAAGGCTCAAAACATAAAGGATGATGTTGTTATCCGGTATAAAGAC 1197
|||||
391 GlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTr 407
|||||
1198 CAAGTAGAGTTTGAAACCATTTGACGGGGAATTTGGTATATTGGAAGAATG 1247
|||||
407 pLysAspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgP 424
|||||
1248 GAAGGATGGTCTGCTCGAACACACATATAAAGGAGTAGTGGTGTTCGAA 1297
|||||
424 heGlnThrSerArgArgValPheLeuValSerProAspSerLeuArgGln 440
|||||
1298 TCCAGACAACAAGAGCTGTATCTCTGTTGGGCCAGATTCTGTAATGCCAG 1347
|||||
441 LeuGlyValGluAspThr 446
|||||
1348 CTTGGAATTCGAATTC 1365
```

seq_name: /SIDS5/gcdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ62625
seq_documentation_block:
ID AAQ62625 standard; cDNA to mRNA: 2557 BP.
XX

```
AC AAQ62625;
XX 18-JAN-1995 (first entry)
XX Rat N-acetyl-glucosamine transferase DNA.
XX production; detection: N-linked asparagine sugar chains;
XX Rat N-acetyl-glucosamine transferase; glycosylation; ds.
XX Rattus sp.
XX Key Location/Qualifiers
XX CDS 158..1501
XX /tag= a
XX /product= N-acetyl-glucosamine transferase
XX JP06113841-A.
XX 26-APR-1994.
XX 08-OCT-1992; 92JP-0270299.
XX 08-OCT-1992; 92JP-0270299.
XX (ASAH ) ASahi Kasei Kogyo KK.
XX WPI; 1994-172739/21.
XX P-PSDB; AARS2657.
XX Rat N-acetyl-glucosamine transferase DNA and protein - useful in
XX prodn. and detection of enzyme
XX Claim 3; Page 10-12; 12pp; Japanese.
XX AAQ62625 encodes a rat N-acetyl-glucosamine transferase (see AARS2657).
XX The enzyme can be used when glycoproteins are produced in yeast so
XX that the natural sugar chain type can be added to the protein of
XX interest, ie. instead of the mannose residues that yeast normally
XX add.
XX Sequence 2557 BP; 535 A; 694 C; 718 G; 610 T; 0 other;
```

alignment_scores:
Quality: 735.50 Length: 422
Ratio: 2.599 Gaps: 13
Percent Similarity: 67.062 Percent Identity: 39.100

alignment_block:
US-09-591-466C-2 x AAQ62625 ..

Align seg 1/1 to: AAQ62625 from: 1 to: 2557

```
30 AlaThrGlnSerGluTyrValAspArgLeuAlaAlaIleGluAlaG 46
|||||
299 GCCAGCTCACCCGTCAGGTCTCCACCTGGCGAGGAGCCCGGGGA 348
|||||
46 uAsnHisCysThrSerGlnThrArgLeuLeuIleAspLysIleSerGlnG 63
|||||
349 G.....TTGGAGCGC 359
|||||
63 InGlnGlyArgValAlaLeuGluGlnMetLysHisGlnAspGln 79
|||||
360 AGCGGGGA.....CTACTGCAGCAGATCAAGGACCATATCT 397
|||||
80 GluCysArgGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyI 96
|||||
398 CTGTGGAGGCAG...AGGTGGAGAGTTCCACCGTGGCCCTCCAGCGTG 444
|||||
96 eLysLysLeuIleGlyAspValGlnMetProValAlaAlaValValM 113
|||||
445 GCCCGGTGTGCTGGGAGCCCTCCACAGCTGTGTATCCCATTCGTGTC 494
|||||
```

Length: 170

```
Ratio: 4.631          Gaps: 0
Percent Similarity: 92.353  Percent Identity: 80.588
alignment_block:
US-09-591-466c-2 x AAX78003  ..
Align seg 1/1 to: AAX78003  from: 1  to: 510

100 lIeGlyAspValGlnMetProValAlaAlaValValMetAlaCysSe 116
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
1  ATCGGAAGCTTGGATCCCGCTGGCGCTGATGTTGTTATGGCTTCAG 50

116 rArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnT 133
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
51  TCGTGCAGACTACTCTGAAGAGGACTGTTAAATCAGTTTAAACATATCAAA 100

133 hrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsn 149
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
101 CTCCGCTGCTTCAAAATATCTCTATTTATATCTCAGGATGATCTCAT 150

150 ProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetG1 166
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
151 CAAGCTGTCAGAGCAAGCTCATTGAGCTATATAATCAATTAACATATATGCA 200

166 hNisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuVala 183
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
201 GCACCTTGATTTTGAACAGCTGCTCAGTGAAGGCCTGGCGAAGTACTG 250

183 laTyrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGlnLeuPhe 199
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
251 CGTACTACAGATGTCAGCTCAGTACAGTGGGCTGGACCCAGTTGTTT 300

200 HisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluI1 216
::: ||||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
301 TACAAGACACAAATTTAGTCGAGTGATATATACTAGAAGATCATATGGAAT 350

216 eAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspa 233
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
351 TGCTCCAGACTCTTTGATTACTTTGAGGCTGCAGCTAGTCTCATGGATA 400

233 rgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArg 249
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
401 GGGATAAAGCAATATAGCTGCTTCATCATGAGTGCATGATAATGCACAGAAG 450

250 GlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheProG1 266
||||| : : : : : ||||||| : : : : : ||||||| : : : : : |||||||
451 CAGTTTGTCATGATCCCTATGCGCTATACCGATCAGATTCTTCCCTGG 500

266 yLeuGlyTrp 269
| |||||
501 CCACGGCTGG 510
```

seq_name: /SID5/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:AAQ25594

seq_documentation_block:

ID AAQ25594 standard; cDNA; 2485 BP.

XX AAQ25594;

XX 03-DEC-1992 (first entry)

XX Rabbit Gnt I cDNA clone.

XX High mannose glycoprotein; UDP-N-acetyl-glucosamine;

KW alpha-3-D-mannoside; beta-1,2-N-acetyl-glucosaminyl transferase I;

KW N-glycans; ss.

OS Oryctolagus cuniculus.

XX Key Location/Qualifiers

XX 50..1393

FT CDS

FT /tag= a

FT /product= Gnt_I

```
FT polyA_signal 2435..2440
XX /*tag= b
PN W09209694-A.
XX 11-JUN-1992.
PD 29-NOV-1991; 91WO-CA00417.
XX 30-NOV-1990; 90US-0620098.
XX (HSCR-) HSC RES & DEV LP.
XX Sarkar M, Schachter H;
XX WPI; 1992-217073/26.
DR P-PSDB; AAR24779.
XX Human and rabbit DNA sequences encoding GNT I enzymes - for
PT converting mannose to hybrid and complex N-glycan(s)
XX Claim 3; Fig 4; 57pp; English.
XX Rabbit liver Gnt I was purified and digested with pepsin, then
CC trypsin. The peptide fragments were sequenced and used to design
CC degenerate PCR primers (see AAQ30220-5). cDNA was prepared from total
CC RNA from rabbit liver. PCR was carried out on the cDNA preparation.
CC One of the two PCR products was cloned into the SmaI site of pGEM7z
CC for sequencing and then used as a riboprobe. The riboprobe was used
CC to screen a rabbit liver cDNA library in lambda gt10. The largest
CC insert in a positive clone was 1.6kb. An 80bp riboprobe was
CC prepared from the 5'-terminal of the 1.6kb insert and used to
CC rescreen the library. The largest cDNA insert was cloned into
CC pGEM-7z to obtain pGEM-7z-rcgntI. The full-length rabbit Gnt I
CC coding sequence was eventually obtained from overlapping clones.
CC See also AAQ25595.
XX
SQ Sequence 2485 BP; 549 A; 623 C; 644 G; 669 T; 0 other;
```

alignment_scores:

Quality: 726.50 Length: 400

Ratio: 2.721 Gaps: 13

Percent Similarity: 66.750 Percent Identity: 39.000

alignment_block:

US-09-591-466c-2 x AAQ25594 ..

Align seg 1/1 to: AAQ25594 from: 1 to: 2485

82 ArgGlnLeuArgAlaLeuValGlnAspLeuGlu.....Se 93

||||| : : : : : ||||||| |||

203 CGTGAAGTATCGCTTAGCTCAGGATGCGAGGTAGAGTTGGAACGTCA 252

93 rLysGlyIleLysLysLeuIleGlyAsp..... 102

||||| : : : : : |||

253 GCGGGGACTGTTGCACGACAGATTAGGAGCACCATGCTCTTTGGAGCCAGC 302

103 ..ValGlnMetProValAlaAla..... 109

||||| : : : : : |||

303 GGTGGAAGGTGCTACTGCAGCCCTCTGCTCAGCCGCATGTGCTGTG 352

110ValValMetAlaCysSerArgTh 118

||||| : : : : : |||||||

353 ACCCCACCCAGCTGTGTATCCCCATCTCTGTAATTGCTGTGACCCAG 402

118 rAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerV 135

||||| : : : : : |||

403 CACC...GTCCCGCCGCTGTTTGGACAGACTACTGCATTATCGCCCTCA. 448

135 a1AlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAsp 151

||||| : : : : : |||||||

449 ..GCTGAGCTGTTCCCGCATTCATTGTGTCAGCCAGCACTGTGGCATGAGGAG 496


```

seq_name: /STD55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT-AAI69837

seq_documentation_block:
ID      AAI69837 standard; cDNA; 1335 BP.
XX
AC      AAI69837;
XX
DT      19-DEC-2001 (first entry)
XX
DE      Human beta-1,2-N-acetylglucosaminyltransferase I cDNA.
XX
KW      Human; beta-1,2-N-acetylglucosaminyltransferase; GnT; GnTI;
KW      maltose-binding protein; MBP; fusion; sugar chain synthesis; ss.
XX
OS      Homo sapiens.
XX
PN      JP2001178453-A.
XX
PD      03-JUL-2001.
XX
PF      24-DEC-1999; 99JP-0368127.
XX
PR      24-DEC-1999; 99JP-0368127.
XX
PA      (TOYM ) TOYOCO KK.
XX      (FUJI/) FUJIYAMA K.
XX
WP      WI; 2001-599905/68.
XX      P-PSDB; AAG67094.
XX
PT      Preparation of maltose-binding protein-N-acetylglucosaminyltransferase
PT      for use in sugar chain synthesis, comprises using Escherichia coli
XX
XX      Claim 5; Page 9-11; 14pp; Japanese.
XX
CC      The invention relates to the preparation of a maltose-binding protein
CC      (MBP)-N-acetylglucosaminyltransferase (GnT) fused protein. The
CC      method comprises transforming Escherichia coli with an expression
CC      vector encoding the fusion protein. GnT antibody is useful in sugar
CC      chain synthesis and in a plasmid pMAL-c2/EcoRI-XbaI fragment to give
CC      a new plasmid pMGNT-1. The present sequence encodes human GnT.
XX
SQ      Sequence 1335 BP; 215 A; 428 C; 441 G; 251 T; 0 other;

alignment_scores:
Quality: 723.50 Length: 396
Ratio: 2.720 Gaps: 13
Percent Similarity: 67.172 Percent Identity: 39.899

alignment_block:
US-09-591-466C-2 x AAI69837 ..

Align seg 1/1 to: AAI69837 from: 1 to: 1335

82 ArgGlnLeuArgAlaLeuValGlnAspLeuGlu.....Se 93
154 CGGGAAGTGGATTCCCTGGCCCAAGACCGAGGTGGAGCTGGAGCGCA 203
93 rLysGlyIleLysLysLeuIleGlyAspVal.....G 104
204 GCGTGGGCTGCTGCAGCAGATCGGGGATGCCCTTCGAGCCAGCGGGGA 253
104 lnMetProValAlaAla..... 109
254 GGGTGGCCACCGGGGCGCTTCGCCGCCACCGCGGTGGCTGTGACCCCC 303
110 .....ValValValMetAlaCysSerArgThrAspTy 120
304 GCGCCGGCGGTGATTCCCATCTCGTGCATCGCTGTGACCGAGCACT.. 351
120 rLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAla 137

```

```
352 .GTTCCGGCGCTCCTGGACAAGCTCTCGATATATCGGCCCTCG...GCTG 397
137 erLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArg 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
398 AGCTCTTCCCATCATCTAGCCAGGACTGCGGCACAGAGACGCC 447
154 LysLeuAlaLeuSerTyrGlyGln...LeuThrTyrMetGlnHisLeuAs 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
448 CAGGCCATCGCTCTACGGCAGCGCTCACGCACATCCGGCAGCCGA 497
169 pTyrGlu.....ProValHisThrGluArgProGlyGluL 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
498 CTGAGCAGCATTTGGGTGCCCGCCGACACCGCAAGTTCCAGGCC.... 543
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTyrAlaLeuAspGln 197
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544 .....TACTACAAGATCGCGGCCACTACCGCTGGCGCTGGGCCAG 585
198 LeuPheHisLysHisAsnPheSerArgValIleLeuGluAspAspMe 214
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
586 GTCITTCGGCAGTTTCGCTCCCGCGCGCTGGTGGAGGATGACCT 635
214 tGluIleAlaAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuL 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
636 GGAGGTGGCCCGCGACTTCTTCGAGTACTTTCGGGCCACCTATCCGCTGC 685
231 euAspArgAspLysSerIleMetAlaIleSerSerTyrAsnAspAsnGly 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
686 TGAAGGCCGACCCCTCCTCTGTGTGGCTCTCGGCTGGAATGACAACGGC 735
248 GlnArgGlnPheVal.....GlnAspProAspAlaLeuTyrArgSerAs 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
736 AAGGAGCAGATGTGTGGACGCGCAGCAGGCTGAGCTGCTTACCGCACCGA 785
262 pPhePheProGlyLeuGlyTyrMetLeuSerLysSerThrTyrSerGluL 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
786 CTTTTTCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 835
279 euSerProLysTyrProLysAlaTyrTrpAspAspTyrLeuArgLeuLys 295
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
836 TGGAGCCCAAGTGGCCAAAGGCTTCTGGGACGACTGGATGGCGGCGCG 885
296 GluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgThrTy 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
886 GAGCAGCGCAGGCGCGGCTGCTGACGCCCTGAGATCTCAAGAAGCAT 935
312 rAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTyrL 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
936 GACCTTTGGCCGCAAGGCTGTGAGCCACGCGGAGTTCTTTGACCGACACC 985
329 euGluProIleLysLeuAsnAspValGlnValAspTyrLysSerMetAsp 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
986 TCAAGTTTATCAAGCTGAACACGAGCTTTGTGCACTTCAACCCAGCTGGAC 1035
346 LeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuVa 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1036 CTCTCTTACCTGACGCGGAGGCGCTATGCCGAGATTTC..... 1074
362 lLysLysAlaLysProIleHisGlyAlaAsp.....AlaValLeuLysA 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1075 .....CTGCCCGCGCTCTACGGTCTCTCCCGAGCTGCAGGTGGAGAAAG 1117
377 laPheAsnIleAsp.....GlyAspValArgIleGlnTyrArg 389
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1118 TGAGGACCAATGACCGAAGGAGCTGGGGAGGTGGGGTGCAGTATACG 1167
390 AspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluG 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1168 GGCAGGACAGCTTCAAGCTTTTCGCCAAGGCTCTGGGTGTATGGATGA 1217
406 uTyrLysAspGlyValProArgAlaAlaTyrLysGlyIleValValPheA 423
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1218 CCTTAAGTCGGGGTTCCGAGAGCTGGCTACCGGGGTATTTGTACCTTCC 1267
```

```
423 rgPheGlnThrSerArgValPheLeuValSerPro 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1268 AGTTCCGG...GGCGCGCTGTCCACCTGGCGGCCCA 1302

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:AAQ25595

seq_documentation_block:
ID   AAQ25595 standard; cDNA; 3230 BP.
AC   AAQ25595;
XX
DT   03-DEC-1992 (first entry)
XX
DE   Human Gnt I cDNA clone.
XX
KW   High mannose glycoprotein; UDP-N-acetyl-glucosamine;
KW   alpha-3-D-mannoside; beta-1,2-N-acetyl-glucosaminyl transferase I;
KW   N-glycans; chronic myeloid leukaemia; ss.
XX
OS   Homo sapiens.
XX
FH   Key Location/Qualifiers
FT   CDS 903..2240
FT     /*tag= a
FT     /product= human_Gnt_I
XX
PN   W09209694-A.
XX
PD   11-JUN-1992.
XX
PF   29-NOV-1991; 91WO-CA00417.
XX
PR   30-NOV-1990; 90US-0620098.
XX
PA   (HSCR-) HSC RES & DEV LP.
XX
PI   Sarkar M, Schachter H;
XX
DR   WPI: 1992-217073/26.
DR   P-PSDB; AAR24781.
XX
XX
XX   Human and rabbit DNA sequences encoding GNT I enzymes - for
XX   converting mannose to hybrid and complex N-glycan(s)
XX
XX   Claim 12; Fig 6; 57pp; English.
XX
XX   A rabbit Gnt I cDNA probe was prepared using PCR on rabbit liver
XX   cDNA. The probe was used to screen an amplified human genomic DNA
XX   library in lambda EMBL3 prepared from chromosomal DNA of chronic
XX   myeloid leukaemia cells. The coding sequence was located in a 4.0kb
XX   fragment of human genomic DNA by screening with a probe containing
XX   the entire coding region of the rabbit Gnt I cDNA. The fragment was
XX   sequenced and found to contain an ORF coding for a protein of 445
XX   amino acids (2 less than the rabbit enzyme). The similarity between
XX   the human and rabbit enzyme coding sequences is 85%.
XX   See also Q25594.
XX
SQ   Sequence 3230 BP; 575 A; 938 C; 913 G; 804 T; 0 other;

alignment_scores:
Quality: 719.50 Length: 396
Ratio: 2.715 Gaps: 13
Percent Similarity: 66.919 Percent Identity: 39.646

alignment_block:
US-09-591-466C-2 x AAQ25595 ..
Align seg 1/1 to: AAQ25595 from: 1 to: 3230

82 ArgGlnLeuArgAlaLeuValGlnAspLeuGlu.....Se 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

1056 CGGGAAGTGATTCCGCTGGCGCCAAAGACCGCGAGGTGGAGCTGGAGCGCAG 1105
93 rLysGlylleLysLysLeuIleGlyAspVal.....G 104
1106 CGGTGGGCTGCTGACGAGATGGGGATGCCCTGTGCGAGCAGCGGGGA 1155
104 InMetProValAlaAa..... 109
1156 GGGTGGCCACCGCGGCCCTCCGCCCCAGCGCGGTGTCCTGTGACCCCC 1205
110ValValMetAlaCysSerArgThrAspTy 120
1206 GCGCCGCGGTGATTCCTGTCATCCCTGTGACCGCAGCACT.. 1253
120 rLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaAs 137
1254 .GTTCGGCGTGGCTGGCAAGCTGTGCATTTATCGCCCTCTC...GCTG 1299
137 erLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspValArg 153
1300 AGCTCTTCCCATCATCTGTTAGCCAGACTGCGGCACGAGGACGGCC 1349
154 LysLeuAlaLeuSerTyrGlyGln....LeuThrTyrMetGlnHisLeuAs 169
1350 CAGCCATCCCTTACGGCAGCGGGTGCACACATCCGCGCAGCCGA 1399
169 pTyrGlu.....ProValHisThrGluArgProGlyGluL 181
1400 CCGTACGAGCAGCATTCGGTGCCTGGCGGACACCGCAAGTTCACGGC.... 1445
181 euValAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGln 197
1446TACTACAAGATCGCGCGCACTACCGCTGGCGCTGGGCCAG 1487
198 LeuPheHisLysHisAsnPheSerArgValIleIleLeuCluAspPhe 214
1488 GTCTTCGGCAGTTTCCTTCCCGCGCGGTGGTGGTGGAGGATGACCT 1537
214 tGluIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuL 231
1538 GGAGTGGCGCGGACTTCTCGAGTACTTCGGGCGCACTATCCGCTGC 1587
231 euAspArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGly 247
1588 TGAAGCCGACCCCTCCCTGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1637
248 GlnArgGlnPheVal.....GlnAspProAspAlaLeuTyrArgSerAs 262
1638 AAGGACGAGATGGTGACGCCAGCAGCGGCTGAGCTCTACCGCACCA 1687
262 pPhePheProGlyLeuGlyTyrMetLeuSerLysSerThrTrpSerGluL 279
1688 CTTTTTCCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 1737
279 euSerProLysTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLys 295
1738 TGGAGCCCAAGTGGCAAGGCTTCTGGGACGACTGGATGGCGGCGG 1787
296 GluAsnHisArgGlyArgGlnPheIleArgProGluValLysArgThrTy 312
1788 GAGCAGCGCAGGGCGGCTGTCATCGCTGAGATCTCAAGAACGAT 1837
312 rAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePheLysGlnTyrL 329
1838 GACCTTTGGCCCAAGGTGTGACACGCGGCGAGTCTTTGACCCAGCAC 1887
329 euGluProIleLysLeuAsnAspValGlnValAspTrpLysSerMetAsp 345
1888 TCAAGTTTATCAAGTGAACACGAGCTTGTGCATCTCACCGACCTGGAC 1937
346 LeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheGlyAspLeuVa 362
1938 CTGCTTACCTGACGGGGGCTATGACCGGAGATTTC..... 1976

362 llLysLysAlaLysProIleHisGlyAlaAsp.....AlaValLeuLysA 377
1977CTGCGCGCGCTCTACGGTGTCCCGCAGCTCAGGTGGAGAAG 2019
377 laPheAsnIleAsp.....GlyAspValArgIleGlnTyrArg 389
2020 TGAGGACCAATGACCGGAGAGCTGGGGAGGTGGCGGTGCATATACG 2069
390 AspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluG 406
2070 GCGAGGACAGCTTCAAGGCTTCGCCCAAGGCTCTGGGTGTATGGATGA 2119
406 uTrpLysAspGlyValProArgAlaalaTyrLysGlyIleValValPheA 423
2120 CCTTAGTCTGGGGTTCGAGAGCTGCTACCGGGGTATTGTCACTTCC 2169
423 rgPheGlnThrSerArgArgValPheLeuValSerPro 435
2170 AGTTCGG...GCCCGCGGTGTCACCTGGCGCCCA 2204
seq_name: /SID5/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.ABL22693
seq_documentation_block:
ID ABL22693 standard; DNA; 2367 BP.
XX ABL22693;
XX DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 19552.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW Pharmacetical; gene; ds.
KW Drosophila melanogaster.
OS
XX WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Claim 1; SEQ ID NO 19552; 2lpp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (AB57737-AB572072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2367 BP; 601 A; 659 C; 581 G; 526 T; 0 other;

```

alignment_scores:
  Quality: 665.00      Length: 333
  Ratio: 2.737        Gaps: 10
  Percent Similarity: 72.973  Percent Identity: 43.544

alignment_block:
  US-09-591-456C-2 x ABL22693  ..

  Align seg 1/1 to: ABL22693  from: 1  to: 2367

106  ProValAlaValValValMetAlaCysSerArgThrAspTyrLeuG1 122
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
896  CCCGTGTTCCAGTCGTAGTCTTCGCCTCAATCGGGTGTGCG..GTGAA 942
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
122  uArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerLysT 139
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
943  GAAGTCATCGATAACTTGGTTCAGTACAGGCCAGCGCTG...GAGCAGT 989
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
139  yrProLeuPheIleSerGlnAspGlySerAsnProAspValArgLysLeu 155
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
990  TCCCATATTGTTGCACAGGACTGGCGGATGAGCCCAAGGAGGCA 1039
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
156  AlaLeuSerTyrGly...GlnLeuThrTyrMetGlnHisLeuAspTyrG1 171
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1040  ATCTCTTCGTATGGCAACAGCTACGCTCATCGAGCAGCTGATCTCAG 1089
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
171  uProValHisThrGluArgProGlyGlu....LeuValAlaTyrTyrL 186
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1090  CGACATC...ACAGTCTGCCGAGGAGAGAAAAATTCAGAGCTACTACA 1138
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
186  ysIleAlaArgHisTyrLysTyrPalaLeuAspGlnLeuPheHisLysH 202
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1137  AGATACCCGACACTACGGCTGGGCCCTGAACACCACTTCT...GCTGTG 1183
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
203  AsnPheSerArgValIleIleLeuGluAspAspMetGluIleAlaAla 219
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1184  GGCTTTGAGTTCGTATATCTCGAGGACGATCTGAACGTGGCGCCGGA 1233
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
219  pPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysS 236
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1234  CTTCCTTTGAGTACTTCCTGGGCAGCCACAAGCTGTCTCAAGCAGGACC 1283
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
236  erIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheVal 252
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1284  GCCTGTGTGCGTGTCCGCGTGAATGCAATGCAAGCTGCTGTGTGTG 1333
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
253  Gln.....AspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1334  GACGGCGCGCAGCCGAGCTGCTACCGCACCGATTTCTTCCCGGTCT 1383
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
267  uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1384  CGCGTGGATGCTCACAAAGATCTGTGGCGGAGCTGCTCGTCAAAATGG 1433
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
284  rOLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAAsnHisAtgG 300
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1434  CCAAAATCTCTGGGATGATGGATACGTCATCCCGCCAGCCCAAGAT 1483
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
301  ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluH 317
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1484  CCGGTGTGCATTAGCCCGGAAATATCACGCACTCGCACGTTTGGAAAA 1533
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
317  sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLys 334
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1534  AGCGCTTTCCAAACGGTTGTCTTCGATAAGTATCTGAAGCACATTAAAC 1583
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
334  euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLe 350
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1584  TGACGGAGGACTTGTGCACTTTACAAAAATCAATATGACGTACCTGCTG 1633
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
351  GlnAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLys 367
      :|||  :|||  :|||  :|||  :|||  :|||  :|||  :|||
1634  AAGGCAATTAACGATAACGCTTCTTCGGCGCGGCTT...TATACGTATCC 1680
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

367 oileHisGlyAlaAspAlaValLeuLys...AlaPheAsnIleAspGlyA 383
||||| |||| ::: :: : : : : : : : : : : : :
1681 CATTTGTACGTAGCATGCAGTCGGCGGAACCTGATTAGAATCGAAGGTC 1730

383 spValArGileGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArg 399
||||| ||||| ||||| ::: : : : : : : : : : : :
1731 CAGTTCGCATTCAATACACTACTAGGACGACGTACAAGCGGACAACTAAG 1780

400 GlnPheGlyIlePheGluGluTrpLysAspGlyValProArgAlaLaTy 416
:
1781 ATGCTGGGTCTTATGGATGATTTCGAAGACGGGTGTCCGCCGAGCTGCCTA 1830

416 rLysGlyIleValValPheArgPheGlnThrSerArgValPheLeu 432
| ||||| ||||| |||| : : : : : : : : : : : :
1831 CCATGGCAGTCGTCTCTTC...TACTACAAACAAACGGCGGTGCATCTG 1876

seq_name: /SID5/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF93806

seq_documentation_block:
ID AAF93806 standard; cDNA; 2250 BP.
XX
AC AAF93806;
XX
DT 23-MAY-2001 (first entry)
XX
DE Human cDNA encoding a membrane or secretory protein clone PSEC0120.
XX
KW Human; secretory protein; membrane protein; vaccine; gene therapy;
KW rheumatoid arthritis; diabetes; ss.
XX
OS Homo sapiens.
XX
PN EP1067182-A2.
XX
PD 10-JAN-2001.
XX
PF 07-JUL-2000; 2000EP-0114090.
PR
PR 08-JUL-1999; 99JP-0194179.
PR 11-JAN-2000; 2000JP-0118775.
PR 02-MAY-2000; 2000JP-01183766.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
XX
DR WPI: 2001-093989/11.
DR P-PSDB; AAB88379.
XX
PT Nucleic acids encoding secretory proteins/membrane proteins, useful in
XX gene therapy or as candidate target molecules in drug development -
XX
FS Claim 1; SEQ ID 125; 609pp + CD ROM; English.
XX
CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
CC which encode human secretory or membrane proteins represented by
CC AAB88317 - AAB88419. Included in the invention are primers
CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the
CC cDNA sequences of the invention. The invention also includes methods for
CC the production of antibodies directed against the proteins, and cDNA
CC sequences, which can be used in vaccines. The polynucleotide sequences
CC can be used in gene therapy. The polynucleotide sequences and the
CC proteins they encode may be used in the prevention, treatment and
CC diagnosis of diseases associated with inappropriate secretory
CC protein/membrane protein expression. The nucleic acids and complementary
CC sequences may also be used as DNA probes in diagnostic assays
CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the
CC presence of similar nucleic acid sequences in samples. They may also be
CC used to study the expression and function of secretory proteins/membrane
CC polypeptides and their role in metabolism. The polypeptides may be used
CC as antigens in the production of antibodies against them and in assays to
CC identify modulators (agonists and antagonists) of expression and

CC agents. The antibodies and antagonists may also be used as therapeutic agents to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the polypeptides in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). Examples of diseases which may be treated include rheumatoid arthritis and diabetes.

XX
SQ Sequence 2250 BP; 408 A; 661 C; 686 G; 495 T; 0 other;

alignment_scores:
Quality: 579.50 Length: 250
Ratio: 3.167 Gaps: 5
Percent Similarity: 73.200 Percent Identity: 46.000

alignment_block:
US-09-591-466C-2 x AAF93806 ..

Align seg 1/1 to: AAF93806 from: 1 to: 2250

194 AlaLeuAspGlnLeuPheHisIysHisAsnPheSerArgValIleLele 210
400 AGCCTCACCAGGCTCTCCGGCAGATTCTCCCGCGCGGTGGTGT 449
210 uGluAspMetGluIleAlaAlaAspPhePheAspTyrPheGluAlag 227
450 GGAGGATGACCTGGAGTGGCCCGGACTCTTCGAGTACTTTCAGGCCA 499
227 lyaAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerTrp 243
500 CCTATCCGCTGCTGAAGCGCGACCCCTCCCTGCTGTGCTTCGCGCTGG 549
244 AsnAspAsnGlyGlnArgGlnPheVal.....GlnAspProAspAlaLe 258
550 AATGACACGCGAAGCAGCAGATGCTGGACGCCAGCAGCCTGACTGCT 599
258 uTyrArgSerAspPheProGlyLeuGlyTrpMetLeuSerLysSert 275
600 CTACCGCACCGACTTTTCCCTGGCCTGGCTGGCTGTGGCGGAGC 649
275 hrTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrp 291
650 TCTGGCTCAGCTGGAGCCCAAGTGGCCAAAGGCTCTCTGGGACGACTGG 699
292 LeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVa 308
700 ATGCGCGCGCGCGGACGACGCGCGCGCGCGCTGCATACGCCCTGAGAT 749
308 lCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPheP 325
750 CTCAGAAGACGATGACCTTGGCCCGCAAGGGTGTGAGCCACGGGCGAGTCT 799
325 heLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrp 341
800 TTGACGACGACCTCAAGTTTATCAAGCTGAACACGACGAGTTGTGCACTTC 849
342 LysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisph 358
850 ACCCAGCTGGACCTGCTTACCTGCACCGGGGAGCCCTATGACCGAGATT 899
358 eGlyAspLeuValLysLysAlaLysProIleHisGlyIleAlaAsp.....A 373
900 C.....CTCGCGCGCGCTCTACGGTGTCTCCCGCGCTGC 931
373 laValLeuLysAlaPheAsnIleAsp.....GlyAspValArg 385
932 AGGTGGACAAGTCAGGACCAATGACCGGAGGAGGTGGGGAGGTTCGG 981
386 lIedInTyrArgAspGlnLeuAspPheGluAspIleAlaArgGlnPheG 402
982 GTGCAGTATACGGGACGGACAGCTTCAAGGCTTTCGCCAAGGCTCTGGG 1031
402 yIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLysGlyI 419

1032 TGTCATGGATGACCTTAAGTCGGGGTTCGAGAGCTGGTACCGGGGTA 1081
419 leValValPheArgPheGlnThrSerArgArgValPheLeuValSerPro 435
1082 TTGTCACCTTCCAGTTCGCG...GGCGCGCTGTCCACCTGGCGGCCCA 1128

seq_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABL22692

seq_documentation_block:
ID ABL22692 standard; DNA; 4721 BP.
XX
AC ABL22692;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19549.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX
PS Claim 1; SEQ ID NO 19549; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4721 BP; 1175 A; 1113 C; 1154 G; 1279 T; 0 other;

alignment_scores:
Quality: 506.00 Length: 453
Ratio: 2.057 Gaps: 20
Percent Similarity: 54.305 Percent Identity: 32.450

alignment_block:
US-09-591-466C-2 x ABL22692/rev ..

Align seg 1/1 to reverse of: ABL22692 from: 1 to: 4721

106 ProValAlaAlaValValMetAlaCysSerArgThrAspTyrLeuGl 122
2826 CCGCGTTTCCAGTCGTAGTCTTCGCTGCAATCGGTGTCG...GTGAA 2780

Wed Aug 14 11:48:26 2002

```

1248 CGAATGGAGAAAGATGAT.....GACTTCACACCTGGACCCAGCTTGC 1291
401 eGlyIlePheGluGluTriPlysAspGlyValProArgAlaAlaTyrLysG 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1292 CAAGTGCCTCCATATCTGGACACCTGGATGTG...CGTGGCAACCATCGG 1338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
418 lylleValValPheArgPheGlnThrSerArgValPheLeuVal... 433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1339 GCCTG.....TGAGATGTGTTTCGGAAGAAGAACCACTTCCTCGGTGGTG 1382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 .....SerProAspSerLeuArgGln 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1383 GGGTCCCGGCTTCCCTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1415
seq_name: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT-AA160103
seq_documentation_block:
ID AA160103 standard; cDNA; 2711 BP.
XX
AC AA160103;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4092.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WPI: 2001-442253/47.
DR P-PSDB; AAM40947.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 4092; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression.
CC

```


CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX

SQ Sequence 2711 BP; 608 A; 755 C; 777 G; 571 T; 0 other;

alignment_scores:
 Quality: 363.00 Length: 361
 Ratio: 1.658 Gaps: 18
 Percent Similarity: 60.565 Percent Identity: 29.917

alignment_block:

US-09-591-466C-2 x AAI60103 ..

Align seg 1/1 to: AAI60103 from: 1 to: 2711

```

103 ValGlnMetProValAlaAlaValValMetAlaCysSerArgThrAs 119
1029 CTCATGTGCTGTGGCT.....GTCAATGCAGGGAAACCGACCCAA 1069
119 pTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrSerValA 136
1070 TTACTCTCAGGATGCTGGGCTCTCTGCTTTCAGCCAGGGGTGTCTC 1119
136 laSerLysTyrProLeuPheIleSerGlnAspGlySerAsnProAspVal 152
1120 CTCAGATGATAACAGATTTCATT.....CAGCGCTACTATGAGGAACCC 1163
153 ArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAs 169
1164 ATGGATGTGGTGGCAGCTGTTTGGT.....CTGAGGGGCATCCA 1201
169 pTyrGluProValHisThrGluArgProGlyGluLeuValAlaTyrTyrL 186
1202 GCATCTCCCATCAGCATCAAGATGCC.....C 1230
186 ysIleAlaArgHisTyrLysTyrTrpAlaLeuAsp.....GlnLeuPhe 199
1231 GCCTGTCTCAGCAGCTACAAGCCAGCCCTCACTGCCACTTTCAACCTGTTT 1280
200 HisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluI 216
1281 CCGAGGCCAAGTTTGCT.....GTGGTTCTGGAAGAGGACCTGGACAT 1324
216 eAlaAlaAspPheAspTyrPheGluAlaGlyAlaThrLeuLeuAspA 233
1325 TGCTGTGGATTTTTCAGTTTCTGAGCCCAATCCACTCACTACTGGAGG 1374
233 rgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArg 249
1375 AGGATGACAGCCGTACTGCATCTCTGCTGGAATGACCCAGGGGTATGAA 1424
250 GlnPheValGlnAspProAspAlaLeuTyrArgSerAspPheProG 266
1425 CACACGGCTGAGGCCACCACTACTGTACCGTGTGGAGACCATGCCTGG 1474
266 yLeuGlyTrpMetLeuSerLysSerThrTrp...SerGluLeuSerProL 282
1475 GCTGGGCTGGGTGCTCAGGAGGCTCTTGTAAGAGGAGGCTTGAGCCCA 1524
282 ysTrpPro.....LysAlaTyr...TrpAspAspTrpLeuArgLeu 294
1525 AGTGGCCTACACCGGAAAGCTCTGGGATTGGACATGTGGATGCGGATG 1574
295 LysGluAsnHisArgGlyArgGlnPheIleArgProGluValCysArgTh 311
1575 CCTGAACAACGCCGGGGCGAGATGTCATCCTCCTGACGTTTCCCGATC 1624
311 rTyrAsnPheGlyGluHisGlySerSerLeu...GlyGlnPhePheLysG 327

```

```

:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1625 CTACCACTTTGGCATCGTCGGCCTCAACATGAATGGCTACTTTTCACGAGG 1674
327 InTyrLeuGluProIleLysLeuAsnAspValGln...ValAspTrpLys 342
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1675 CCTACTTCAGAAGCACAAGTTCACACAGGTTCACAGGTTCACAGCTCAGG 1724
343 SerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPheG 359
:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1725 AATGTGGAC.....AGTCTGAAGAAAGAGCTTATGAAGTGGAGTTCA 1768
359 yAspLeuValLysLysAlaLysProIle...HisGlyAlaAspAlaVal 375
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1769 CAGGCTGCTCAGTGAGGCTGAGGTTCTGGACCACACGAAGAACCCTTGTG 1818
375 euLysAlaPheAsnIleAspGlyAsp.....Val 384
:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1819 AAGACTCTTTCTCCGCCAGACACAGAGGCCACACCTACGTGGCCTTTATT 1868
385 ArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGlnPh 401
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1869 CGAATGGAGAAAGATGAT.....GACTTCACCACCTGGACCCAGCTTGC 1912
401 eGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLysG 418
:||||:||||| ||| :||| :||| ||| ||| :||| :|||
1913 CAAGTGCCTCCATATCTGGACCTGGATGTG...CGTGGCAACCATCGGG 1959
418 yIleValValPheArgPheGlnThrSerArgArgValPheLeuVal... 433
||||:||||| ||| :||| :||| ||| ||| :||| :|||
1960 GCCTG.....TGGAGATTGTTTCGAAGAAGAACCCACTTCTCTGGTGGTG 2003
434 .....SerProAspSerLeuArgGln 440
||||| ||||| :||| :||| :||| :||| :|||
2004 GGGGTCCCGGCTTCCCCCTACTCAGTGAAGAAG 2036

```

OM of: US-09-591-466C-2 to: GenEmbl:* out_format : pfs
Date: Aug 13, 2002 10:11 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+p2n,model -DEV=xlp
-O=/cpn2.1/USPTO_spool/US09591466/runat_12082002_095312.4260/app_query.fasta_1.511
-DB=GenEmbl -OFMT=fastap -SUFFI=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pls -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=2000000000 -USER=US09591466.@CEN1.1.7593
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:
Query: US-09-591-466C-2
Query length: 446
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 2163.780000

score_list:	Strd Orig	zScore	EScore	Len	Documentation
Sequence					
gb_pl:STU249878	+ 2345.00	3793.95	5.8e-203	1641	! A249878 Solanum tuberosum mRNA
gb_pat:AS95053	+ 2345.00	3793.78	5.9e-203	1669	! AS95053 Sequence 1 from Patent
gb_pl:STU249879	+ 2219.00	3588.86	1.5e-191	1691	! A249879 Solanum tuberosum pat
gb_pl:NTY16832	+ 2201.00	3557.76	8.3e-190	2031	! Y16832 Nicotiana tabacum mRNA
gb_pl:NTA249883	+ 2198.00	3554.63	1.2e-189	1708	! A249883 Nicotiana tabacum mRN
gb_pat:AS95055	+ 2198.00	3554.63	1.3e-189	1737	! AS95055 Sequence 3 from Patent
gb_pl:NTA249882	+ 2197.00	3552.27	1.7e-188	1836	! A249882 Nicotiana tabacum mRN
gb_pl:STU249880	+ 1948.00	3149.63	4.5e-167	1499	! A249880 Solanum tuberosum mRN
gb_pl:ATH243198	+ 1781.00	2876.25	7.6e-152	1820	! A243198 Arabidopsis thaliana
gb_pat:ATH249881	+ 1768.00	2855.07	1.1e-150	1830	! A249881 Arabidopsis thaliana
gb_pat:AS95057	+ 1768.00	2854.94	1.2e-150	1854	! AS95057 Sequence 5 from Patent
gb_pl:ATF20D10	- 1133.50	1779.59	9.2e-91	147215	! AL035538 Arabidopsis thalian
gb_pl:ATCHRIV89	- 1133.50	1776.51	1.4e-90	199789	! AL161593 Arabidopsis thalian
gb_pl:ATF22113	- 1131.50	1780.89	7.8e-91	93760	! AL035539 Arabidopsis thalian
gb_hcg:AF004082	- 886.00	1379.35	1.8e-68	120538	! AP004082 Oryza sativa chromo
gb_ro:RATNAGT	+ 735.50	1173.65	5.2e-57	2546	! D16302 Rat mRNA for N-acetylgl
gb_pat:E07296	+ 735.50	1173.61	5.2e-57	2557	! E07296 cDNA encoding rat liver
gb_ov:XLGNTI	+ 735.50	1172.99	5.7e-57	2508	! Y16819 Xenopus laevis mRNA for
gb_to:AF087456	+ 732.50	1175.22	4.3e-57	1344	! AF087456 Mesocricetus auratus
gb_to:BC008629	+ 732.50	1170.95	7.4e-57	1894	! M73491 Mouse N-acetylglucosam
gb_to:MUSTRANX	+ 732.00	1167.56	1.1e-56	2651	! BC006629 Mus musculus, Similar
gb_hcg:AL060829	+ 732.00	1167.39	1.2e-56	2695	! L07037 Mus musculus (clone E31
gb_om:RABUDPNA	+ 732.00	1121.02	4.4e-54	267628	! AL606829 Mus musculus chromo
gb_to:AC057192	+ 726.50	1159.27	3.3e-56	2485	! M57301 O. cuniculus UDP-N-acety
gb_to:CGU65794	+ 725.50	1163.84	1.8e-56	1344	! AF087457 Mesocricetus auratus
gb_to:AF343963	+ 725.50	1155.40	1.8e-56	1344	! AF057194 Homo sapiens cDNA FL
gb_to:CG65791	+ 723.50	1160.59	2.8e-56	1344	! AF343963 Cricetus griseus muta
gb_pr:HUMGLCNAC	+ 723.50	1160.59	2.8e-56	1344	! AF343963 Cricetus griseus N
gb_pr:AC022413	+ 723.50	1153.93	6.5e-56	2602	! M55621 Human N-acetylglucosam
gb_to:AF343964	+ 723.50	1153.62	6.8e-56	2683	! BC003575 Homo sapiens, mannosi
gb_hg:AC023255	+ 723.50	1147.96	1.4e-55	4705	! M61829 Human alpha-1,3-mannosy
gb_to:AF343965	- 723.50	1110.30	1.4e-53	166525	! AC022413 Homo sapiens chromo
gb_ov:XLGLCNAC	+ 718.50	1152.46	7.9e-56	1344	! AF343964 Cricetus griseus mu
gb_in:AF082010	+ 701.00	1147.59	1.5e-55	1344	! AF343965 Cricetus griseus mu
gb_in:AF082010	+ 701.00	1147.59	1.5e-55	1344	! AF343965 Cricetus griseus mu
gb_in:AF082010	+ 643.00	1054.76	2.2e-50	2828	! AF251495 Drosophila melanogast
gb_in:AF082010	+ 643.00	1030.53	4.9e-49	1719	! AF082010 Caenorhabditis elega
gb_in:AF082012	+ 633.00	1013.67	4.2e-48	1322	! AF082012 Caenorhabditis elega

gb_pat:AX136203 + 579.50 921.36 5.9e-43 2250 ! AX136203 Sequence 125 from
gb_in:AF082011 + 550.00 874.69 2.3e-40 1982 ! AF082011 Caenorhabditis ele
gb_hg:AC014285 - 506.00 767.87 2.1e-34 65730 ! AC014285 Drosophila melano
gb_in:AC009218 - 508.00 759.10 6.4e-34 156753 ! AC009218 Drosophila melan
gb_in:AC008347 - 506.00 758.56 6.9e-34 165519 ! AC008347 Drosophila melan
seq_name: gb_pl:STU249878
seq_documentation_block:
LOCUS STU249878 1641 bp mRNA linear PLN 04-JAN-2002
DEFINITION Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI
gene), clone Al.
ACCESSION AJ249878
VERSION AJ249878.1 GI:18076139
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (sites)
AUTHORS Wenderoth,I., Tjaden,J. and von Schaewen,A.
TITLE Isolation and characterization of different plant
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional
analyses in the Arabidopsis cgl mutant, and in potato and tobacco
antitense plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1641)
AUTHORS von Schaewen,A.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,
GERMANY
FEATURES
source Location/Qualifiers
1..1641
/organism="Solanum tuberosum"
/variety="Desiree"
/db_xref="taxon:4113"
/clone="Al"
/tissue_type="leaf mesophyll"
/note="sink"
39..1379
/gene="GntI"
39..1379
/gene="GntI"
/EC_number="2.4.1.101"
/function="initiates complex-type glycans in the Golgi"
/codon_start=1
/evidence=experimental
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAC80697.1"
/db_xref="GI:18076140"
/translation="MGNKFCFLRLVLLVVAALFYIOMRLFATOSEYVDRLAAATE
AENHVSQTRLLLDKISQGRYVALEQMKHQDQCRLRALVQDLSEKGIKKLIGD
VOMPVAASVLMACSRDYLERIKTSIKYVSKYPLFISODGSDNPVVKLALSYG
QLTYGHLEDPVHTPELVAIYKIRHYKALDQHPKHNFSLVILDDMEIAA
DFYDFEAGATLLDRKSIAMISWSDNGORQVODPDALYRSDFFPGLGWLKSTW
SDSPKPKAYNDWILRLKENHGRQFIPEVCRTYNGEHSGLGDFQYLEPIKL
NLSQVDMKSDLSYLLLEDNYKHFGLVKKAKPIHGADAVLKAFLNIDGVDVRYQDL
DVEDIARQGFGEFGEKDGVPVRAIKGVIFVFQTSRRVFLVSPDSLRQLGVEDT"
BASE COUNT 485 a 302 c 377 g 477 t
ORIGIN
alignment_scores
Quality: 2345.00 Length: 446
Ratio: 5.258 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-591-466C-2 x STU249878 ..
Align seg 1/1 to: STU249878 from: 1 to: 1641

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl 17
|||||
39 ATCAGAGGGAACAAGTTTTCGTTGATTACGGTACCTTCGTCGTGGC 88
|||||
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
89 TGCCTCGGCTTCATCATACATACAGATGGGCTTTTTCGGCAGACAGTCAG 138
|||||
34 LuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
139 AATATGTAGACCCCTTGCTGTCGAATTCAGCAGAAAAATCATTTGACA 188
|||||
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
189 AGTCAGACCAGATTGCTTATTGACAAGATTAGCCAGCAGCAAGAAAGAT 238
|||||
67 lValAlaLeuGluGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
239 AGTAGCTCTTTGAAGAACAAATGAAGCATCAGGACCAGGAGTGCCGCAAT 288
|||||
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
289 TAAGGCTCTTGTTCAGGATCTTTGAAGTAAGGCATAAAAAAGTTAATC 338
|||||
101 GlyAspValGlnMetProValAlaAlaValValValMetAlaCysSerAr 117
|||||
339 GGAGATGTCAGATGCCAGTGGCAGCTGTAGTTGTTATGGCTTGCAGTCG 388
|||||
117 qThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
389 TACTGACTACCTCGAGAGGACTATTAAATCCATCTTAAATAACCAACAT 438
|||||
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
439 CTGTGTCATCAAAATATCTCTTTTCATATCCAGGATGGATCAAAATCCT 488
|||||
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
489 GATGTAAGAAAGCTTCTTGACCTATGTCAGCTGACGTATATGCGACA 538
|||||
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
539 CTTGGATTATGAACCTGTGCATACTGAAAGACCAGGGAACCTGTTGCAT 588
|||||
184 yTyrLysIleAlaArgHisTyrLysTyrGlyGlnLeuAspGlnLeuPheHis 200
|||||
589 ACTACAAGATTGCACGTCATTAACAAGTGGCATTTGGATCAGCTGTTTAC 638
|||||
201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
639 AAGCAATAATTTAGCCGCTGTATCATACTAGAAAGATCATATGCAAAATGC 688
|||||
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArg 234
|||||
689 TGCTGATATTTTGGACTATTTTGGAGCTGGAGCTACTCTTCTTTCAGACAG 738
|||||
234 sPlySerIleMetAlaIleSerSerTyrAsnAspAsnGlyGlnArgGln 250
|||||
739 ACAAGTCGATTATGGCTATTTCTTCTTGGAAATGCAATGCACAAAGGCAG 788
|||||
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
|||||
789 TTCGTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTCCTGCTCT 838
|||||
267 uGlyTyrMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284
|||||
839 TGGATGATGCTTTCAAAATCAACTTGGTCCGAACATATCTTCCAAAAGTGC 888
|||||
284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
|||||
889 CAAAGGCTTACTGGGATGACTGGCTAAGGCTGAAGCAAAAAATCACAGAGT 938
|||||
```

```
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317
|||||
939 CGACAATTTATTCGCCCAGAAAGTTTGCAGAACGCTACAAATTTTGTGTAGCA 988
|||||
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
|||||
989 TGGTCTTAGTTTGGGCAGTTTTTTTAAAGCAGTATCTTTGAGCCCAATTAAGC 1038
|||||
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
|||||
1039 TAAATGATGTCAGAGTTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1088
|||||
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
|||||
1089 GAGGACAACTATGTGAACACTTTGGCGACTTGGTTAAAAAGGCTAAGCC 1138
|||||
367 oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
|||||
1139 CATCCAGGAGCTGATGCTGTTTGAAGCATTTAACAATAGATGTCATG 1188
|||||
384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
|||||
1189 TGGGTATTTCAGTACAGAGACCAACTAGACTTTGAAGATATCGCTCGACAG 1238
|||||
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
|||||
1239 TTTGGCATTTTGAAGAAATGAAGGATGGTGTACCCAGCGCAGCATATAA 1288
|||||
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuVals 434
|||||
1289 AGGATAGTAGTTTTCGGGTTTCAAAACATCTAGACGTGTGTTCTCTGTTT 1338
|||||
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
|||||
1339 CCCTGATCTCTTCGACAACTTGGAGTTGAAGATACT 1376
|||||
seq_name: gb_pat:A95053
seq_documentation_block:
LOCUS A95053 1669 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 1 from Patent WO9929879.
ACCESSION A95053
VERSION A95053.1 GI:6779205
KEYWORDS .
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 1669)
AUTHORS Von.S.A.
TITLE VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)
ACTIVITY
Patent: WO 9929879-A 1 17-JUN-1999;
VON SCHAEWEN ANTJE (DE)
FEATURES
Location/Qualifiers
source
1..1669
/organism="Solanum tuberosum"
/strain="DESIREE"
/db_xref="taxon:4113"
/cell_type="BLATTZELLEN"
/tissue_type="MESOPHYLL"
/clone_lib="LAMBDA ZAP II (ECORI)"
/dev_stage="(B) CLON(E): GNTI-AL(K)"
152
53..1393
/gene="CGL"
53..1393
/gene="CGL"
/standard_name="GNTI"
/EC_number="2.4.1.101"
```

5'UTR
gene
CDS

```

/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/notes="ERSTE GNTI-SEQUENZ AUS KARTOFFEL (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE 1"
/db_xref="GI:6779206"
/translation="MRGNFCFDRLRYLLVVAALAFIYIOMRLFATQSEYVVDRLAAAE
AENHCTSQRLLDIKISQOQGRVLALEQMKHQDQCRQLRALVDLESKGIKKLIGD
QMPVAAVVMACSRDYLERTIKSLKYQTSVASKYPLIFISQGSNPDVRKLAISYG
QLTMOHLDYEPVHTYRPGELVAYYKIAHYKWLDPDLFKHNFSRVILLEDDEMTAA
DFDYPEAGATLLDRDKSITMAISSNNDNGOROFVODPDALYRSDPEFPGILGWLSKSTW
SELSPKWPKAYDDMLRKLKENHGRQFIRPEVCRTYNFGEHSSGLQCFKQXLEPKL
NDVQVWKSMDSLLELYEDNYKHFGLVKKAPITGADAVLKAFFNIDGVRIOYRDQL
DFDIARQGIFFEENKDGVPRAAYKGIIVFRFQTSRRVFLVSPDSLRLQIGVED"
<80. .>139
CDS
/genes="CGL"
/standard_name="MEMBRANANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMBRANANKER (AMINO ACID 10-29)"
/notes="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70463.1"
/db_xref="GI:6779207"
/translation="LRYLLVVAALAFIYIOMRLF"
1394. .1655
3'UTR 489 a 312 C 387 g 481 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 2345.00      Length: 446
  Ratio: 5.258          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-591-466C-2 x A95053  ..

Align seg 1/1 to: A95053 from: 1 to: 1669

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
53 ATGAGAGGGACAGATTTCCTTGATTACGGTACCTTCCTCGTGGGCG 102

17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
103 TGCTCTCGCCTTCATCTACATACAGATGCGGCTTTTCGCGACACAGTCAG 152

34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
153 AATATGTAGACCGCTTGCTGCTGCAATTGAAGCAGAAAAATCATTTGACA 202

51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
203 AGTCAGACACAGATTGCTATTATGACAGATTAGCCACGACGAGAGAGT 252

67 lValAlaLeuGluGluClnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
253 AGTAGCTCTTTGAAGAACAAATGAAGCATCAGGACACGAGAGTGCCTGCAAT 302

84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
303 TAAGGGCTCTTGTTTCAGGATCTTGAAGTTAAGGCATATAAAAGTTAATC 352

101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117
|||||
353 GGAGATGTGCAGATGCCAGTGGCAGCTGTAGTTGTTATGGCTTGCAGTCG 402

117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
403 TACTGACTACCTGGAGAGGACTATTAAATCCATCTTAAATAACCAACAT 452

```

```

134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
453 CTGTTTGATCAAAATATCTCTTTTCATATCCAGAGTGGATCAAAATCCT 502

151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHI 167
|||||
503 GATGTAGAAAGCTTGGCTTTTGAGCTATGGTCAGCTGATATATGCAGCA 552

167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
553 CTTCGATTATGAACCTGTGCATCTAGAAAGACGAGGGAACCTGGTTGCAT 602

184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
|||||
603 ACTACAAAGATTGCACGCTTATACAAATGGGCATTTGGATCGACTCTTTCAC 652

201 LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAl 217
|||||
653 AAGCAATATTTTACCCGTGTTTATCATCTAGAAATGATATGGAAATGTC- 702

217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
|||||
703 TGCTGATTTTTTTGACTATTTTGGAGGCTGGAGCTACTCTTCTTGACAGAG 752

234 spLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
|||||
753 ACAAGTCGATTATGGCTATTTCTTCTTGGAAATGACAATGGACAAGGCAG 802

251 PheValGlnAspProAspAlaLeuTyrArgSerAspPhePheProGlyLe 267
|||||
803 TTCGTCCAAGATCCTGATGCTCTTTACCGCTCAGACTTTTTTCTCGGTCT 852

267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpp 284
|||||
853 TGGATGGATGCTTTTCAAAATCAACTTGGTCCGAACATATCTCCAAGTGC 902

284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
|||||
903 CAAGGCTTACTGGGATGACTGGCTAAGGCTGAAGAGAAAATACAGAGGT 952

301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHI 317
|||||
953 CGACAAATTTATTCGCCAGAAAGTTTCAGAAACGTACAATTTTGGTGAGCA 1002

317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
|||||
1003 TGGTTCCTAGTTTGGGCGAGTTTTTTAAGCAGTATCTTGAGCCAAATTAAGC 1052

334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
|||||
1053 TAAATGATGTCCAGGTGATTGGAAGTCAATGGACCTAAGTTACCTTTTG 1102

351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
|||||
1103 GAGGACAACTATGTGAACACACTTTGGCGACTTGGTTTAAAAAGGCTAAGCC 1152

367 otleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
|||||
1153 CATCCAGCGAGCTGATGCTGTTTGAAGCATTTAACATAGATGGTGATG 1202

384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
|||||
1203 TCGGTATTTCAGTACAGAGACCAACTAGACTTTTGAAGATATCGCTCGACAG 1252

401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
|||||
1253 TTTGGCATTTTGAAGATGGAAGGATGGTGACCACGGGCGAGCATATAA 1302

417 sGlyIleValValPheArgPheGlnThrSerArgValPheLeuValS 434
|||||
1303 AGGGATAGTAGTTTTTCCGGTTTCAAAACATCTAGACGTGTGTTCTCTT 1352

```

434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
 1353 CCCCTGATTCTCTCGACAACCTTGAGTTGAAGATACT 1390
 seq_name: gb_pl:STU249879

seq_documentation_block:
 LOCUS STU249879 1691 bp mRNA linear PLN 04-JAN-2002
 DEFINITION Solanum tuberosum partial mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A6.

ACCESSION AJ249879
 VERSION AJ249879.1 GI:18076141
 KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.
 SOURCE potato.
 ORGANISM Solanum tuberosum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.

REFERENCE 1 (sites)
 AUTHORS Wenderoth, I., Tjaden, J. and von Schaewen, A.

TITLE Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1691)
 AUTHORS von Schaewen, A.

TITLE Direct Submission
 JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY

FEATURES
 source

1..1691
 /organism="Solanum tuberosum"
 /variety="Desiree"
 /db_xref="taxon:4113"
 /clone="A6"
 /tissue_type="tuber"
 /note="sink"
 1..1298
 /gene="GntI"
 1..1298
 /gene="GntI"
 /EC_number="2.4.1.101"
 /function="initiates complex-type glycans in the Golgi"
 /codon_start=3
 /product="N-acetylglucosaminyltransferase I"
 /protein_id="CAC80698.1"
 /db_xref="GI:18076142"

/translation="VAALAFIYIQMLFATOSEYVDRILAAIEAEDHCTQSRLIDK
 ISQOGRVLEEOMKQDQECRLALVODLESKGIKKLIGDVQMPVAAVVMACSR
 ADYLEKTKISLKYOTSVAKYPLFIQDGNPDVRKLALSYDQLYMQHLDPEVHT
 ERPELVAYYKIAHYKWDLPFKHNFSSRIILEDDMETAADFFDYFEAGATLLDR
 DKSIMAISSMNDGROFVQDPYALYRSDFPPLGLWMLSKSTWSELSPKWPAYWDDW
 LRLKENHRRGROFIRPEVCRTYFNGEHGSSLGQFFKQVLEPIKLDNDVQDWKSMDSYL
 LEDNAVYHFGDLVAKPKPIHGADAVLAFNIDGVRQYRQDLDFDIARQFIFEW
 KDGVPRAAYKGIYVFRQTSRRVFLVAPDSLRQLGVEDT"

BASE COUNT 518 a 307 c 381 g 485 t
 ORIGIN

alignment_scores:
 Quality: 2219.00 Length: 431
 Ratio: 5.172 Gaps: 0
 Percent Similarity: 99.536 Percent Identity: 98.144

alignment_block:

US-09-591-466C-2 x STU249879 ..

Align seg 1/1 to: STU249879 from: 1 to: 1691

16 valAlaAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrG1 32
 ||||||||||||||||||||||||||||||||||||||||||||||||||||

3 GTGGCTGCTCTCGCCTTCATCTACATACAGATCGCGCTTTTCGGGACACA 52
 32 nSerGluTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisC 49
 |||||||||||||||||||||||||||||||||||||||||||||||||
 53 GTCAGAATAATGTAGACCCCTTGCTGCTGCAATTTGAAGCAGCAAGATCAT 102
 49 yThrSerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGly 65
 |||||||||||||||||||||||||||||||||||||||||||||||||
 103 GTACAAGTCAGACAGATTCCTTATTGACAAGATTATGCCAGCAGCAAGGA 152
 66 ArgValValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysAr 82
 |||||||||||||||||||||||||||||||||||||||||||||||||
 153 AGAGTAGTGGCTCTCTGAAGAACAATGAAGCTCAGGACCAGGAGTGCCG 202
 82 gGlnLeuArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysL 99
 |||||||||||||||||||||||||||||||||||||||||||||||||
 203 GCAATTAAGGGCTCTTGTTCCAGGATCTTTGAAGTAAGGGCATAAAAAGT 252
 99 eutLeGlyAspValGlnMetProValAlaAlaValValMetAlaCys 115
 |||||||||||||||||||||||||||||||||||||||||||||||||
 253 TAATCGGAGATGTCAGATGCCAGTGGCGCTGTAGTTGTTATGGCTTGC 302
 116 SerArgThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrG1 132
 |||||||||||||||||||||||||||||||||||||||||||||||||
 303 AGTCGTGCTGACTACCTGGAGAAGACTATTAAATCCATCTTAAATACCA 352
 132 nThrSerValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerA 149
 |||||||||||||||||||||||||||||||||||||||||||||||||
 353 AACATCTGTTGCATCAAAATATCCTCTTTTCATATCCCAGGATGGATCAA 402
 149 snProAspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMet 165
 |||||||||||||||||||||||||||||||||||||||||||||||||
 403 ATCCTGATGTTAGAAAGCTTGCTTTGAGCTATGATCAGCTGACGTATATG 452
 166 GlnHisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuVa 182
 |||||||||||||||||||||||||||||||||||||||||||||||||
 453 CAGCACTTGGATTTTGAACCTGTGCATACTGAAAGACCAGGGCAACTGGT 502
 182 lAlaTyrTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuP 199
 |||||||||||||||||||||||||||||||||||||||||||||||||
 503 TGCATACTACAAGATTGCAGCTATTACAAGTGGGCACTTGATCAGCTGT 552
 199 heHisLysHisAsnPheSerArgValIleIleLeuGluAspAspMetGlu 215
 |||||||||||||||||||||||||||||||||||||||||||||||||
 553 TTCACAAGCATATATTTAGCCGTGTTATCATACTAGAAGATGATATGGAA 602
 216 lIleAlaAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAs 232
 |||||||||||||||||||||||||||||||||||||||||||||||||
 603 ATTGCTGCTGATTTTGTGACTATTTTGAGGCTGGAGCTACTCTTCTTGA 652
 232 pArgAspLysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnA 249
 |||||||||||||||||||||||||||||||||||||||||||||||||
 653 CAGAGATAAGTCGATTATGGCTATTTTCTTCTTGGAAATGACAAATGGACAAA 702
 249 rgGlnPheValGlnAspProAspAlaLeuTyrArgSerAspPhePro 265
 |||||||||||||||||||||||||||||||||||||||||||||||||
 703 GGCAGTTCGTCCAAAGATCTTATGCTCTTTACCGCTCAGACTTTTTCCT 752
 266 GlyLeuGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLy 282
 |||||||||||||||||||||||||||||||||||||||||||||||||
 753 GGTCTTGGATGGATGCTTTCAAAATCAACTTGTGCGAACTATCTCCAAA 802
 282 sTrpProLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisA 299
 |||||||||||||||||||||||||||||||||||||||||||||||||
 803 GTGGCCAAAGGCTTACTGGGATGACTGGCTAAGGCTCAAAGAAATACACA 852
 299 rgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGly 315
 |||||||||||||||||||||||||||||||||||||||||||||||||
 853 GAGGTGCGAAATTTATTCGCCCAGAGTTTGCAGAACATACAAATTTTGGT 902
 316 GluHisGlySerSerLeuGlnPhePheLysGlnTyrLeuGluProI1 332
 |||||||||||||||||||||||||||||||||||||||||||||||||
 903 GAGCATGTTCTTAGTTTGGGCGAGTTTTTTTAAGCAGATATCTTGAGCCAA 952

332 eLysLeuAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrL 349
|||||
953 TAAGCTAAATGATGTCCAGGTTGATTGGAAGTCAATGGACCTAAGTTACC 1002
349 euLeuGluAspAsnTyrValLysHisPheGlyAspLeuValLysAla 365
|||||
1003 TTTTGGAGGACAACTATGTGAACACTTTGGCGACTTGGTTAAAGGCT 1052
366 LysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspG 382
|||||
1053 AAGCCCATCCACGCTGCTATGCTGTTTGAAGACATTTAAACATAGATGG 1102
382 YAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAla 399
|||||
1103 TGATGTCGCTATTTCAGTACAGACCACTAGACTTTGAAGATATCGCTC 1152
399 rgGlnPheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAla 415
|||||
1153 GGCAGTTTGGCATTTTGAAGAATGGAAGATGGTGTACCACGGCAGCA 1202
416 TyrLysGlyIleValValPheArgPheGlnThrSerArgArgValPheLe 432
|||||
1203 TATAAAGGGATAGTAGTTTCGGGTTTCAACATCTAGACGTGTGTCTCT 1252
432 uValSerProAspSerLeuArgGlnLeuGlyValGluAspThr 446
|||||
1253 TGPTGCCCTGATTCTCTCGACAACCTTGAGTTGAAGATACG 1295

seq_name: gb_pl:NTY16832

seq_documentation_block:
LOCUS NTY16832 2031 bp mRNA linear PLN 24-AUG-1999
DEFINITION Nicotiana tabacum mRNA for beta-1,2-N-acetylglucosaminyltransferase, complete CDS.
ACCESSION Y16832
VERSION Y16832.1 GI:5764048
KEYWORDS beta-1,2-N-acetylglucosaminyltransferase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 2031)
Strasser,R., Mucha,J., Schwiela,H., Altmann,F., Glossl,J. and
Steinkellner,H.
Molecular cloning and characterization of cDNA coding for beta1,
2N-acetylglucosaminyltransferase I (GlcNAc-TI) from nicotiana
tabacum
Glycobiology 9 (8), 779-785 (1999)
9935389
2 (bases 1 to 2031)
Steinkellner,H.
Direct Submission
Submitted (06-MAR-1998) H. Steinkellner, Zentrum fuer Angewandte
Genetik, BOKU-Wien, Muthgasse 18, 1190 Wien, AUSTRIA
Location/Qualifiers
1. .2031
/organism="Nicotiana tabacum"
/cultivar="xanthi"
/db_xref="taxon:4097"
1. .76
77. .1417
/EC_number="2.4.1.101"
/codon_start=1
/product="alpha-1,3-mannosyl-glycoprotein
beta-1,2-N-acetylglucosaminyltransferase"
/protein_id="CAB53347.1"
/db_xref="GI:5764048"
/translation="MRGYKPCDFRYLLILAAVAFIYIOMLFATQSEYADRLAAALE
AENHKTSOTPLLDQISQOGRIVALEQMKRQDCRQLRALVQDLESKGIKRLIGN
VQMPVAAYVMACNRADYLEKTKISLKQISVAPKYPPIFISQDGHDPVKKILALSYD
QLTYMHLDFEPVHTERPGEIJAYYKIARHYKWLQDLFYKHNFSRVIILEDDEIAP

5'UTR
CDS

DDFFFEAGATLLDRDKSTMAISSHNDNCQMDFVQDPYALYRSDFFPGLGWLMSKSTW
DELSPWPKAYMDDWLRLKENHGRQFIRPEYCRSTNFGEHSSLGQFFKQYLEPIKL
NDVQVDMKSMDSLYLLEDNYVKHAKPIHGADAVLAKAFNIDGVRIOYRDOL
DFEDTARQFGIFPEWKGVPRAAYKGIWVFRYQTSRRVLVGPDSLIQULQUNEDT*
1418. .2031
3'UTR 606 a 380 c 427 g 618 t
BASE COUNT
ORIGIN
alignment_scores:
Quality: 2201.00 Length: 446
Ratio: 5.037 Gaps: 0
Percent Similarity: 97.982 Percent Identity: 93.049
alignment_block:
US-09-591-466C-2 x NTY16832 ..
Align seg 1/1 to: NTY16832 from: 1 to: 2031
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValValAl 17
|||||
77 ATCAGAGGGGTACAAGTTTGTCTGTGATTTCGGGTACCTCTCATCTTGGC 126
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
127 TGTGTGGCTTCTCATCTACATACAGATCGCGCTTTTTCGCACACAGTCAG 176
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
177 AATATGCAGATCGCTTGTCTGCAATTCAAGCAGAAATCACTGTACA 226
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
|||||
227 AGTCAGACCAGATGTGCTTATTGACCAGATTAGCCAGCAGCAAGGAAGAT 276
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
|||||
277 AGTGTCTCTTGAAGACAAATGAAGCGTCAGACACGAGGATGCGCAGCT 326
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
|||||
327 TAAGGCTCTTGTTCAGGATCTTGAAGTAAGGCGATAAAAAAGTTGATC 376
101 GlyAspValGlnMetProValAlaAlaValValValMetAlaCysSerAr 117
|||||
377 GGAATGTACAGATGCCAGTGGCTGCTGTAGTTGTTATGGCTTGCATCG 426
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS 134
|||||
427 GGCTGACTACTGGAAAAAGACTATTAAATCCATCTTAAAAATACCAATAT 476
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
|||||
477 CTGTTGCGCAAAATATCTCTTTTCATATCCAGGATGGATCACAATCCT 526
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
|||||
527 GATGTTAGGAAGCTTGTCTTGTAGCTATGATCAGCTGACGTATATGCAGCA 576
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
|||||
577 CTTGGATTTGAACCTGTGCATACCTGAAAGACCGAGGGAGCTGATTGCAT 626
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
|||||
627 ACTACAAAATTCAGCTCATTAACAAGTGGCATTTGGATCAGCTGTTTAC 676
201 LysHisAsnPheSerArgValIleIleLeuGluAspMetCulileAl 217
|||||
677 AAGCATAAATTTAGCCGCTTATCATCTACTAGAAAGATGATATGAAATTCG 726
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA 234
| |||||

727 CCTGATTTTTTTGACTTTTTTTGAGCTGGAGCTACTCTCTTGACAGAG 776
234 spLysserileMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
|||||
777 ACAAGTCGATATGGCTATCTCTTGAAGTACAAATGGACAAATGCAG 826
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
|||||
827 TTTGTCCAAGATCTTATGCTCTTTACCGCTCTGATTTTTTCCCGTCT 876
267 uGlyTrpMetLeuSerLySerThrTrpSerGluLeuSerProLystrp 284
|||||
877 TGGATGGATGCTTTCAAAATCTACTTTGGGACGAATATCTCCAAATGGC 926
284 roLyssAlaTyrTrpAspAspTrpLeuArgLeuLyssGluAsnHisArgGly 300
|||||
927 CAAAGCTTACTGGGACGACTGGCTAAGACTCAAGAGAATCACAGAGT 976
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317
|||||
977 CGACAATTTATTCGCCCAGAAAGTTTCAGATCATATAATTTTGGTGAGCA 1026
317 sGlySerSerLeuGlyGlnPhePheLyssGlnTyrLeuGluProIleLyssL 334
|||||
1027 TGGTCTAGTTTGGGCGAGTTTTCAGGAGTATCTTTGAGCCAAATTAAC 1076
334 euAsnAspValGlnValAspTrpLyssSerMetAspLeuSerTyrLeuLeu 350
|||||
1077 TAAATGATGTCCAGGTTGATTGGAAGTCAATGGACCTTAGTTACCTTTTG 1126
351 GluAspAsnTyrValLyssHisPheGlyAspLeuValLyssLyssAlaLyssr 367
|||||
1127 GAGGACAATTACGTGAACACCTTTGGTGACTTTGGTTAAAAAGGCTAAGCC 1176
367 oIleHisGlyAlaAspAlaValLeuLyssAlaPheAsnIleAspGlyAspV 384
|||||
1177 CATCCATGGAGCTGATGCTGTTTGAAGCATTTAACATAGATGGTGATG 1226
384 aIArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
|||||
1227 TGGGTATTAGTACAGATCACTAGACTTTGAAGATATCGCACGGCAA 1276
401 PheGlyIlePheGluGlnTrpLyssAspGlyValProArgAlaAlaTyrLy 417
|||||
1277 TTTGGCATTTTGAAGAAATGGAAGGATGGTGATCCACGGGCGCATATAA 1326
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuValIs 434
:::|||||
1327 AGCAATAGTGGTTTCCGGTACCAACGTCACAGCGTGATTCCTTGTTG 1376
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
:::|||||
1377 GCCTGATTGCTTCAACAACTCGGAAATGAAGATACT 1414

seq_name: gb_pi:NTA249883

seq_documentation_block:
LOCUS NTA249883 1708 bp mRNA linear PLN 04-JAN-2002
DEFINITION Nicotiana tabacum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A9.
ACCESSION AJ249883
VERSION AJ249883.1 GI:18076147
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (sites)
Wenderoth, I., Tjaden, J. and von Schaeuwen, A.
Isolation and characterization of different plant
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional
analyses in the Arabidopsis cgl mutant, and in potato and tobacco

antisense plants
Unpublished
REFERENCE 2 (bases 1 to 1708)
AUTHORS von Schaeuwen, A.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) von Schaeuwen A., Pflanzenphysiologie,
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,
GERMANY
FEATURES
source
1..1708
/organism="Nicotiana tabacum"
/variety="Samsun NN"
/db_xref="taxon:4097"
/clone="A9"
/tissue_type="leaf mesophyll"
/note="sink"
113..1453
/gene="GntI"
113..1453
/gene="GntI"
/EC_number="2.4.1.101"
/function="initiates complex-type glycans in the Golgi"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAC80702.1"
/db_xref="GI:18076148"
/translation="MRGNKFCDFRYLLILAAVFIYTMRLFATQSEYADRLAAAE
AENHCTQTRLLIDQISLQOGRIVALEQMKRQDECRQLRALVODLESKGIKKLIGN
VQMPVAAVYNACNRADYLETKISILKYQISVASKYPLFIQDGSHPDVRKLALSYD
QLTYMQHLDPFVHTERPGELIATYKIAHHYKAWLDQLFYKHNFSRVIILEDDEIAP
DFFDFEAGATLLDRDKSIMAISNNNDGOMQFVDDPYALYRSDFPGLGWLMSKSTW
DELSKPWPKAYWDDMLRLENHGRQFIRPEVCRITYNFGHSSGLGQFFKOYLEPKL
NDVQDWKSMDSLYLLEDNYVKGFDLKGAKPIHGADAVLKAFNIDGVRIOYRDQL
DFENARQFIFEEMKDGVPRAAYKGIIVFRYQTSRRRVFLVGHDSLQQLGIET"
BASE COUNT 506 a 326 c 380 g 496 t
ORIGIN
alignment_scores:
Quality: 2198.00 Length: 446
Ratio: 5.030 Gaps: 0
Percent Similarity: 97.982 Percent Identity: 92.825
alignment_block:
US-09-591-466C-2 x NTA249883 ..
Align seg 1/1 to: NTA249883 from: 1 to: 1708
1 MetArgGlyAsnLyssPheCysPheAspLeuArgTyrLeuLeuValVal 17
|||||
113 ATGAGAGGGAACAAGTTTTCGTGTGATTTCCGGTACCTCTCATCTTGGC 162
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
|||||
163 TGCCTGTCGCTTCATCTACACAGATGCGGCTTTTTCGACACACATCAG 212
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
|||||
213 AATATGCAGATCGCTTCGTGTCGTAATTAAGCAGCAAAAATCATCTGACA 262
51 SerGlnThrArgLeuLeuLeaspLyssIleSerGlnGlnGlnGlyArgVa 67
|||||
263 AGCCAGACAGATTTGCTTATTGACCAGATTAGCCTGCAGCAAGGAAGAT 312
67 lValAlaLeuGluGlnMetLyssHisGlnAspGlnGlnCysArgGlnL 84
:|||||
313 AGTTGCTCTTGAGAACAAATGAAGCGTCAGGACAGGAGTGCCGCAAT 362
84 euArgAlaLeuValGlnAspLeuGluSerLyssGlyIleLyssLyssLeuIle 100
|||||
363 TAAGGCTCTTCTTTCAGGATCTTGAAGTAAGGGCATAAAAAAGTTGATC 412
101 GlyAspValGlnMetProValAlaAlaValValMetAlaCysSerAr 117

413	GGAANTGTACAGATCCGAGTGGCTGTGTAGTTGTATTATGCCTTGCAATCG	467
117	gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThrS	134
463	GGCTGATTACCCTGGAAGAAGACTATTAAATCCATCTTTAAATAACCAATAT	512
134	erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro	150
513	CTGTTCGGTCAAATAATPCCTCTTTTCATATCCAGATGGATCACAATCCT	562
151	AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetClnHi	167
563	GATGTCAGGAAGCTTCGTTTGACCTATGATCACGTGACGTATATCGACGA	612
167	sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAlaIat	184
613	CTTGGGATTTTGAACCTGTGCATCTGAAAGACGAGGGAGCTGATTCGAT	662
184	yTrYrLysIleAlaArqHisTyrLysTrpAlaLeuAspGlnLeuPheHis	200
663	ACTACAAAATTCACGCTCATTAACAGTGGCCATTTGGATCAGCTGTTTTATC	712
201	LysHisAsnPheSerArgValIleIleLeuGluAspAspMetGluIleAI	217
713	AAGCATTAATTTTAGCCGCTTATCATACTAGAAGATGATATGGAAATTCG	762
217	aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArgA	234
763	CCTCGATTTTTTGACCTTTTTTGAGGCTGGAGCTACTCTCTTGACAGAG	812
234	sPlySerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln	250
813	ACAAGTCGATTAATGGCTATTTCTTCTTGGAAATGACAATGGACAAATGCAG	862
251	PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe	267
863	TTTGTCCAAGATCTTATGCTCTTTTACCGCTCAGATTTTTTTCGCGGTCT	912
267	uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrpp	284
913	TGCATGGATGCTTCCAAATCTACTTGGAGCAATTAATCTCAAAGTGGC	962
284	rOLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly	300
963	CAAAGGCTTACTGGGACACTGGCTAAGACTCAAAGAGAAATCACAGAGT	1012
301	ArgGlnPheIleArqProGluValCysArgThrTyrAsnPheGlyGluHi	317
1013	CGACAATTTATCGCCCCAGASTTTGCAGAACATATAATTTTGGTGAGCA	1062
317	sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL	334
1063	TGGTTCATAGTTTGGGCGAGTTTTTCAAGCAGTACTTTTGAGCCAATTAAC	1112
334	euanAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu	350
1113	TAAATGATGCCAGTTGATTGGNAGTCAATGGACCTTAGTTACCTTTTG	1162
351	GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr	367
1163	GAGCACAAATTAAGTGAACACACTTTGGTGACTTCGTTAAAAAGCGCTAAGCC	1212
367	oIleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV	384
1213	CATCCATGGAGCTGATGCTGCTTGAAAGCATTTTAACATAGATGGTGATG	1262
384	alArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln	400
1263	TGGGTATTTCAGTACAGAGATCAACTAGACTTTGAAAAATATCCACGGCAA	1312
401	PheGlyIlePheGluGlnTrpLysAspGlyValProArgAlaAlaTyrIby	417

```

1313 TTTGGCATTTTTGAAGATGAAGGATGGTGTACCACGTCGACGATATAA 1362

417 sGlyIleValIleValPheArgPheGlnTrSerArgValPheLeuValS 434
|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1363 AGGAATAGTAGTTTTCGGGTACCAACAGCTCAGACGTTGATTCTCTGTTG 1412

434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
:: |||||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
1413 GCCATGATTCGCTTCAACAACTCGGAATTCGAAGATACT 1450

seq_name: gb_pat:A95055
seq_documentation_block:
LOCUS A95055 1737 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 3 from Patent WO9929879.
ACCESSION A95055
VERSION A95055.1 GI:6779208
KEYWORDS .
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 1737)
Von,S.A.
VEGETABLE GntI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH
A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GntI)
ACTIVITY
Patent: WO 9929879-A 3 17-JUN-1999;
Von SCHAEWEN ANTJE (DE)
FEATURES
source 1..1737
/organism="Nicotiana tabacum"
/strain="SAMSON NN"
/db_xref="taxon:4097"
/cell_type="BLATTZELLEN"
/tissue_type="MESOPHYLL"
/clone_lib="LAMBDA ZAP II (ECORI)"
/clone_lib="(B) CLON(E): GNTI-A9(T)"
/dev_stage="M: SINK' ORGAN"
15..126
127..1467
127..1467
/gene="CGL"
127..1467
/gene="CGL"
/standard_name="GNTI"
/EC_number="2.4.1.101"
/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET.
GLYKOPROTEINEN"
/note="ERSTE GNTI-SEQUENZ AUS TABAK (UNPUBLIZIERT)"
/codon_start=1
/evidence=experimental
/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"
/protein_id="CAB70464.1"
/db_xref="GI:6779209"
/translation="MRGNKCCDFRYLLILAAVFIYTMRLFATQSEYADRLAAAEI
AENHCTSTRLIDQISQQQFVALVEEQMKRQDQECRLRALVQDLESKGIKKLSDY
VOMPYAAVWVACNADRYLERTIKSLIKYQISVASKYPLFIISQDGHDPVKRLASVD
QLTVMHQHLDFFPEVTRPECELLIAYVKIARHWALDQLEFKHNFSEVILDDMEIATP
DFFDFEAGATLLDRKSTIMALSSNDNCQMFQVDPYALYSDFFPGLGWLMSKSTW
DELSPKWPKAYWDDMLRKLKRNKHGQFIRPEVCRITNFGEHSSLGQFFKQWYLEPIKL
NDVQVDWKSMDLSLLEDDNYVKNHFGDLVKKAKPIHGADAVLKAFNIDGVDRIQYRDLQ
DFENIARQFGIFEEMKQDVPRAAYKGIIVFYQTSRRVFLVGHDSLQQLGIEDT"
154..>213
/gene="CGL"
/standard_name="MEMBRANKER EINES TYP II GOLGI-PROTEINS"
/function="MEMBRANKER (AMINO ACID 10-29)"
/note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN
IDENTIFIZIERT"
/codon_start=1
/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"
/protein_id="CAB70465.1"
/db_xref="GI:6779210"
CDS

```



```
3'UTR /translation="FRYLILAAVAFIYTMRLF"
BASE COUNT 1468 . .1723
ORIGIN 510 a 337 c 390 g 500 t

alignment_scores:
  Quality: 2198.00 Length: 446
  Ratio: 5.030 Gaps: 0
Percent Similarity: 97.982 Percent Identity: 92.825

alignment_block:
US-09-591-466C-2 x A95055 ..
Align seg 1/1 to: A95055 from: 1 to: 1737
```

```
1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuValAla 17
127 ATGAGAGGAACAAGTTTGTGTGTTTCCGGTACCTCCCTCATCTTGGC 176
17 aAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
177 TGCTGTGCGCTTCATCTACACACAGATCGCGCTTTTTCGCACACAGTCAG 226
34 luTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
227 ANTATGCAGATCGCTTCTGCTGCAATTTGAAGCAGAAAATCATTTGTACA 276
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlyArgVa 67
277 AGCCAGACCAGATTGCTTATTGACCAGATTAGCCTGCAGCAAGCAAGCAAT 326
67 lValAlaLeuGluGlnMetLysHisGlnAspGlnGlnGlyCysArgGln 84
327 AGTTGCTCTTGAAGCAACAATGAAGCGTCAGGACCGAGGAGTGGCGCAAT 376
84 euArgAlaLeuValIleGlnAspLeuGluSerLysGlyIleLysLysLeu 100
377 TAAGGCGCTTGTTCAGGACTCTTGAAAGTAAGGCAATAAAGTTGATC 426
101 GlyAspValGlnMetProValAlaAlaValIleValMetAlaCysSerAr 117
427 GGAATGTACAGATGCCAGTGGCTGTGTAGTTGTTATGGCTTGCATCG 476
117 gThrAspTyrLeuGluArgThrIleLysSerIleLeuLysTyrGlnThr 134
477 GGCTGATTACCTCGAAAGACTATTAAATCCATCTTAAATATCAAAATAT 526
134 erValAlaSerLysTyrProLeuPheIleSerGlnAspGlySerAsnPro 150
527 CTGTTCGCTCAAAATATCTCTTTTCATATCCAGGATGCCAGTGCATCCT 576
151 AspValArgLysLeuAlaLeuSerTyrGlyGlnLeuThrTyrMetGlnHi 167
577 GATGTCAGGAAGCTGCTTTGAGCTATGATCAGCTACGCTATATGCAGCA 626
167 sLeuAspTyrGluProValHisThrGluArgProGlyGluLeuValAla 184
627 CTTGGATTTTGAACCTGTGCATACTCAAAAGACCAGGGAGCTGATTGCAT 676
184 yTyrLysIleAlaArgHisTyrLysTrpAlaLeuAspGlnLeuPheHis 200
677 ACTACAAAATTTGCAGCTCATATCAAGTGGGCATTTGGATCAGCTGTTTAC 726
201 LysHisAsnPheSerArgValIleIleLeuGluAspMetGluIleAl 217
727 AAGCAATAATTTAGCGCTGTTATCATACTAGAAATGATATGCAATTCG 776
217 aAlaAspPhePheAspTyrPheGluAlaGlyAlaThrLeuLeuAspArg 234
777 CCTGATTTTTTGACTTTTTTTCAGGCTGGAGCTACTCTTCTTGACAGAG 826
234 sPlysSerIleMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGln 250
```

```
||||| 827 ACAAGTCGATTATGCGCTATTCTTCTTGGAAATGACAATGGACAAATGCAG 876
251 PheValGlnAspProAspAlaLeuTyrArgSerAspPheProGlyLe 267
877 TTTGTCCAAGATCCTTATGCTCTTTACCGCTCAGATTTTTCCTCCGGTCT 926
267 uGlyTrpMetLeuSerLysSerThrTrpSerGluLeuSerProLysTrp 284
927 TGGATGGATGCTTTCAAAATCTACTTGGGACGAATTTATCCAAAGTGGC 976
284 roLysAlaTyrTrpAspAspTrpLeuArgLeuLysGluAsnHisArgGly 300
977 CAAAGGCTTACTGGGACGACTGGCTAAGACTCAAGAGAGATCACAGAGT 1026
301 ArgGlnPheIleArgProGluValCysArgThrTyrAsnPheGlyGluHi 317
1027 CGACAATTTATTCGCCCAGAGTTTGCAGAACATATAATTTTGTGTAGCA 1076
317 sGlySerSerLeuGlyGlnPhePheLysGlnTyrLeuGluProIleLysL 334
1077 TGGTTCATGTTTGGGCGAGTTTTCAGACAGTATCTTGAGCCCAATTAAC 1126
334 euAsnAspValGlnValAspTrpLysSerMetAspLeuSerTyrLeuLeu 350
1127 TAAATGATGTCAGGTTGATTGGAAGTCAATGGACCTTAGCTTACCTTTTG 1176
351 GluAspAsnTyrValLysHisPheGlyAspLeuValLysLysAlaLysPr 367
1177 GAGGACAAATTCGTGAACACTTTGGTGACTTGGTTAAAAGGCTAAGCC 1226
367 olleHisGlyAlaAspAlaValLeuLysAlaPheAsnIleAspGlyAspV 384
1227 CATCCATGGAGCTGATGCTGCTTGAAGCATTTAACATAGATGGTGATG 1276
384 aArgIleGlnTyrArgAspGlnLeuAspPheGluAspIleAlaArgGln 400
1277 TGGTATTTCAGTACAGAGATCAACTAGACTTTGAAAATATTCGCACGCCAA 1326
401 PheGlyIlePheGluGluTrpLysAspGlyValProArgAlaAlaTyrLy 417
1327 TTTGGCATTTTGAAGAATGGAAGGATGGTGATACCAGCTGCACATATAA 1376
417 sGlyIleValValPheArgPheGlnThrSerArgArgValPheLeuVal 434
1377 AGGAATAGTAGTTTTTCCGTTACCAAAACGTCACAGCGTGTATCTCTGTTG 1426
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
1427 GCCATGATTGCTTCAACAACCTCGGAATTTGAAGATACT 1464
```

seq_name: gb_pl:NTA249882

seq_documentation_block:
LOCUS NTA249882 1836 bp mRNA linear PLN 04-JAN-2002
DEFINITION Nicotiana tabacum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A4.
ACCESSION AJ249882
VERSION AJ249882.1 GI:18076145
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.

REFERENCE
AUTHORS Wenderoth, I., Tjaden, J. and von Schaewen, A.
TITLE Isolation and characterization of different plant N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1836)

AUTHORS von Schaewen, A.
TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY

FEATURES
SOURCE

Accession	Source	Location/Qualifiers
1	1836	organism="Nicotiana tabacum"
		variety="Samsun NN"
		db_xref="taxon:4097"
		clone="A4"
		tissue_type="leaf mesophyll"
		note="sink"
gene	27	1367
		gene="GntI"
CDS	27	1367

gene

```

gene
27. .1367 /note= "Sink"
CDS
27 1367 /gene="GntI"

```

2

gene- GIII	27	1367
CDS		

```

/gene= Gm1
/EC_number=2.4.1.101"
/function="initiates complex-type glycans in the Golgi"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAC80701.1"
/db_xref="GI:18076146"
/translation="MRGYKCCDFVLLLLAAVAFYIOMRLFATOSEYADRLAAAE
AENHCTSQTRLLIDQISQQGRIVALEEEMKQDQCRLRALVQLESKGTKLIGN
VQMPAAVVMCAVNRHYLEKTIKLYOISVAPKPLFIQSDGSHVQKIALSYD
QITYMHLDFMCPHRTPEGLIAYYKIAHYKWLQQLFYKHNFSRVIILEDMETAP
DFFDFEAGATLLDRDKSLMYSNNNDNQMFQVQPYALYSDFPGLGWLMSKSTW
DELSPKPQAYDWDMLKLNENHGRQFIRPECVSRFYEGHSSLSGQFTKOYLEPKL
NDVQVQSMDSLYLLEDNRYVHFGDLVKKAPIGHCADAVLKAFNIDGVRIOYRDL
DFEDLARQFIFEENQDGPRAAYGVYFRTSRRFLVGPDSLQQLGNET"
536 a 354 c 403 g 543 t
BASE COUNT
ORIGIN

```

alignment scores:

Quality:	2197.00	Length:	446
Ratio:	5.027	Gaps:	0
Percent Similarity:	97.982	Percent Identity:	92.825

alignment_block:

US-09-591-466C-2 x NTA249882

Align seg 1/1 to: NTA249882 from: 1 to: 1836

```

1 MetArgGlyAsnLysPheCysPheAspLeuArgTyrLeuLeuValValAl
17
27 ATGAGAGGTCACAAAGTTTGCTGTGATTTCCGGTACCTCCTCATCTTGGC 76
17 aaAlaLeuAlaPheIleTyrIleGlnMetArgLeuPheAlaThrGlnSerG 34
77 TCGTCTGCCCTTCATCTACATACAGATCGCGCTTTTTCGCACACAGCTCAG 126
34 IuTyrValAspArgLeuAlaAlaIleGluAlaGluAsnHisCysThr 50
127 AATATGCAGATCGCTTGCTGCTCAATTGAAGCAGAAATCACTGTACA 176
51 SerGlnThrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgVa 67
177 AGTCAGACGAGATTGCTATTGTACACGATTAGCAGCAGCAGCAAGGAAGAA 226
67 IValAlaLeuGluGluGlnMetLysHisGlnAspGlnGluCysArgGlnL 84
227 AGTTGCTCTTCAAGACACAAATGAAGCGCTGAGGACCAAGGAGTGGCGACAG 276
84 euArgAlaLeuValGlnAspLeuGluSerLysGlyIleLysLysLeuIle 100
277 TAAGGGCTCTTGTTCAGGATCTTGAAGTAAGGCGATATAAAAGATTGATC 326
101 GlyAspValGlnMetProValAlaIaValValIaMetAlaCysSerAr 117
327 GGAATGTACAGATGCCAGTGGCTGTGTAGTTGTATTGCTGTGCAATCG 376

```

[illegible]

|||||
1277 AGGAATAGTGGTTTCGGTACCAACGCGCCAGCGTATTCCCTGGTTG 1326
434 erProAspSerLeuArgGlnLeuGlyValGluAspThr 446
:|||||
1327 GCCCTGATTCCGTTCAACAACCTCGGAATGAGATACT 1364
seq_name: gb_pl:STU249880

seq_documentation_block:
LOCUS STU249880 1499 bp mRNA linear PLN 04-JAN-2002
DEFINITION Solanum tuberosum mRNA for N-acetylglucosaminyltransferase I (GntI gene), clone A8.
ACCESSION AJ249880
VERSION AJ249880.1 GI:18076143
KEYWORDS GntI gene; N-acetylglucosaminyltransferase I.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (sites)
Wenderoth, I., Tjaden, J. and von Schaewen, A.
Isolation and characterization of different plant
N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional analyses in the Arabidopsis cgl mutant, and in potato and tobacco antisense plants
Unpublished
2 (bases 1 to 1499)
von Schaewen, A.
Direct Submission
Submitted (28-SEP-1999) by Schaewen A., Pflanzenphysiologie, Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck, GERMANY

FEATURES
Source Location/Qualifiers
1..1499
/organism="Solanum tuberosum"
/variety="Desiree"
/db_xref="taxon:4113"
/clone="A8"
/tissue_type="tuber"
/note="sink"
37..1158
/gene="GntI"
37..1158
/gene="GntI"
/EC_number="2.4.1.101"
/function="initiates complex-type glycans in the Golgi"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAC80699.1"
/db_xref="GI:18076144"
/translation="MKRQDQECRLRALVODLESKGIKKLIGDVQMPVAVVVMACSR
ADYLERTIKSLKYQTSVASKYPLFTSQDSNPDVRKLLASLDLYDQLTVMQHLDFEPVHT
ERPGLVAYVKIARHYKWLADLFHKHNSRVITLEDDEMEIAADFTDYFAGATLLDR
DKSLJMAISSNDNCORQFVQDPDALYNSDFPGLGWLMSKSTWSELSKPWKAYWDDW
LRUKENHRGQFLHPEVCRTYFEGHSSLGQFFKQYLEP IKLNDVQVQWKSMDLSYL
LEDNYRHKFDGLVKKAKPHICADAVLKAFNIDGVDVRIQVRLDQDFEAIARQGIFFEW
KGVPRAAAYKGIWVFRFQTSRRVFLVAPDSLRLQGVEDT"
BASE COUNT 452 a 266 c 339 g 442 t
ORIGIN

alignment_scores:
Quality: 1948.00 Length: 374
Ratio: 5.237 Gaps: 0
Percent Similarity: 99.465 Percent Identity: 98.128

alignment_block:
US-09-591-466C-2 x STU249880 ..

Align seg 1/1 to: STU249880 from: 1 to: 1499

73 GlnMetLysHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValG1 89
|||||
34 CAATGAAGCGCTCAGGACCAGGAGTGC CGCAATTAAAGGCTCTTGTTCA 83
89 nAspLeuGluSerLysGlyIleLeuLysLysLeuIleGlyAspValGlnMetP 106
|||||
84 GGATCTTGAAGTAAAGGCATAAAAAGTTAAATCGGAGATGTCAGATGC 133
106 roValAlaAlaValValMetAlaCysSerArgThrAspTyrLeuGlu 122
|||||
134 CAGTGGCAGCTGTAGTTGTTATGCTTGCAGTCGCTGCTACTACTGGAG 183
123 ArgThrIleLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTy 139
|||||
184 AGGACTATTAAATCCATCTTAAATAACCAACATCTCTTGCATCAAAATA 233
139 rProLeuPheIleSerGlnAspGlySerAsnProAspValArgLysLeuA 156
|||||
234 TCCTCTTTTCATATCCAGGATGGATCAAAATCCTGATGTAAGAAAGCTTG 283
156 laLeuSerTyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluPro 172
|||||
284 CTTTGAGCTATGATCAGCTGACGTATATGACGACCTTGGATTTGAACCT 333
173 ValHisThrGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaAr 189
|||||
334 GTCCATACTGAAAGACACAGGGGAACCTGGTTGCATACTACAAGATTGCACG 383
189 gHisTyrLysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerA 206
|||||
384 TCATTACAAGTGGGCTTGGATCAGCTGTTTCACAAGCATATTTTAGCC 433
206 rgValIleIleLeuGluAspAspMetGluIleAlaAlaAspPheAsp 222
|||||
434 GTGTTATCATATAGAGATGATATGGAATTCGCTGCTGATTTTTCAC 483
223 TyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAl 239
484 TATTTTGAGGCTGGAGCTACTCTCTTGACAGACAAAGTCGATTATGGC 533
239 aIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProA 256
|||||
534 TATTTCTTCTTGGAATGCAATGGACAAAGGCAGTTCGTCGAAGATCCTG 583
256 spAlaLeuTyrArgSerAspPheProGlyLeuGlyTrpMetLeuSer 272
584 ATGCTCTTTACCGCTCAGACTTTTTCCTGGTCTTGGATGATGCTTTCA 633
273 LysSerThrTrpSerGluLeuSerProLysTrpProLysAlaTyrTipAs 289
634 AAATCAACTTGGTCGGAACCTATCTCCAAAGTGCCAAAGGCTTACTGGGA 683
289 pAspTrpLeuArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgP 306
684 TGACTGGCTAAGGCTGAAGAAAAATCACAGAGTCCGACAAATTTATTCA 733
306 roGluValCysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGly 322
734 CAGAAGTTTGCAGAACCTACAAATTTTGGTGAGCATGCTCTAGTTGGGG 783
323 GlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnVa 339
784 CAATTTTTAAGCAGTATCTTTAGCCCAATTAAGCTAAATGATGCCAGGT 833
339 lAspTrpLysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValL 356
834 TGATTGGAAGCTCAATGGACCTAAGTTACCTTTTGGAGGACAACTATGTGA 883
356 yHisPheGlyAspLeuValLysLysAlaLysProIleHisGlyAlaAsp 372
884 AACACTTTGGCGACTTGGTTAAAGAGCTTAAGCCCATCCACGAGCTGAT 933
373 AlaValLeuLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrAr 389

```

|||||
934 GCTGTTTGAAGCATTTTAAACATAGATGGTATGCGTATTTCAGTACAG 983
|||||
389 gAspGlnLeuAspPheGluAspIleAlaArgGlnPheGlyIlePheGluG 406
|||||
984 AGACCAACTAGACTTTGAAGCTATCGCTCGACAGTTTGGCATTTTGAAG 1033
|||||
406 luTrpLysAspGlyValProArgAlaAlaTyrLysGlyIleValValPhe 422
|||||
1034 AATGAAGATGGTGTACACCGGCGAGCATATAAAGGGATAGTAGTTTC 1083
|||||
423 ArgPheGlnThrSerArgValPheLeuValSerProAspSerLeuArg 439
|||||
1084 CGGTTTCAAAACATCTAGACGTGTCTCTCTGTTGCCCTGATTCTCTTCG 1133
|||||
439 gClnGlyValGluAspThr 446
|||||
1134 ACAACTGGAGTTGAAGATACT 1155

seq_name: gb_pl:ATH243198
seq_documentation_block:
LOCUS ATH243198 1820 bp mRNA linear PLN 24-AUG-1999
DEFINITION Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I
(GlcNAcT-I gene).
ACCESSION AJ243198
VERSION AJ243198.1 GI:5139334
KEYWORDS GlcNAcT-I gene; glycosyl transferase;
N-acetylglucosaminyltransferase I.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS Bakker,H., Lommen,A., Jordi,W., Stiekema,W. and Bosch,D.
TITLE An arabidopsis thaliana cDNA complements the
N-acetylglucosaminyltransferase I deficiency of CHO lec1 cells
JOURNAL Biochem. Biophys. Res. Commun. 261 (3), 829-832 (1999)
MEDLINE 99373163
REFERENCE 2 (bases 1 to 1820)
AUTHORS Bakker,H.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Bakker H., Molecular Biology, Centre for
Plant Breeding and Reproduction Research, P.O. Box 16 Wageningen,
6700AA, NETHERLANDS
COMMENT Related sequence AL035538.
FEATURES
Location/Qualifiers
1..1820
/organism="Arabidopsis thaliana"
/variety="Colombia"
/db_xref="taxon:3702"
/clone_lib="pABE-AtGNI"
/clone_lib="pABE-A.t.selle"
31..1365
/gene="GlcNAcT-I"
31..1365
/gene="GlcNAcT-I"
/EC_number="2.4.1.101"
/function="glycosyltransferase"
/codon_start=1
/product="N-acetylglucosaminyltransferase I"
/protein_id="CAB45521.1"
/db_xref="GI:5139335"
/translation="MARISCDLRFLLIPAAFMFIYIQMLFQTSQYADRLSSAIESE
NHCTSQMRGLIDEVSIKQSDRLVDELMKNDDEELVOLKDIOTFEKKGIKALTQGGQ
MPVAAYVMAACSRADYLVETVKSIVTYQTVPVASKYPLFIQSDGDAVKSKSLSYNL
TYMQLHDFEVPVTERPGLTAYYKIAHYKWLDDLPYKHKFSRVITLEDDEIAPDF
FDYFEAAASLMDRDKTITMAASSNDNCKQKOFVHDYALYISDFPPLGLWMLKSTWDE
LSPKWPAYWDDWLRLKENHKGQFIRPEVCRTYNFGEHSSLGQFESQYLEPIKLN
VTVDWKAQDLGILTEGNYTKYFSGLVQRAPIQGSDLVLAQNIKDDVRIYKDOVEF
ERTAGEFGIEEWKWDGVPRTAYKGVVVFIQTTRRVFLVGPDSVMQLGIRNS"
532 a 342 c 421 g 525 t
BASE COUNT

```

ORIGIN

alignment_scores:

Quality: 1781.00 Length: 438

Ratio: 4.441 Gaps: 0

Percent Similarity: 91.553 Percent Identity: 75.571

alignment_block:

US-09-591-466C-2 x ATH243198 ..

Align seg 1/1 to: ATH243198 from: 1 to: 1820

```

9 AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleG1 25
|||||
49 GACTTGAGATTCTCTCATCCCGCAGCTTTCATGTTCACTACATCCA 98
|||||
25 nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAa 42
|||||
99 GATGAGGCTTTTCCAGACGCAATCACAGTATGCAGATGCAGTCAGTCCG 148
|||||
42 laileGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
|||||
149 CTATCGAATCTGAGAACCATTCAGTCAATGCGAGGCGCTCATAGAT 198
|||||
59 LysIleSerGlnGlnGlnArgValAlaLeuGluGlnMetLy 75
|||||
199 GAAGTTAGCATCAACAGTCGCGATTGTTGCCCTCGAAGATATGAAGAA 248
|||||
75 sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
|||||
249 CCGCCAGCAGCAAGAACTTGTGAGCTTAAGGATCTAATCCAGACGTTG 298
|||||
92 luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAla 108
|||||
299 AAAAAAAGAAATAGCAAACTCACTCAAGGTGACAGATGCCTGTGGCT 348
|||||
109 AlavaIvalValMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
|||||
349 GCITGATGTTATGGCTCGAGTCGAGTCTGAGACTATCTTGAAGGACTGT 398
|||||
125 eLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeup 142
|||||
399 TAAATCAGTTTTAACATATCAAACTCCCGTTGCTTCAAAATATCCTCTAT 448
|||||
142 heIleSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSer 158
|||||
449 TTATATCTCAGGATGATCTGTCAAGCTGTCAAGACCAAGTCATTGAGC 498
|||||
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHisTh 175
|||||
499 TATAATCAATTAACTATATGCAGCACTTGGATTTGAACCCAGTGTCTAC 548
|||||
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrL 192
|||||
549 TGAAGGCCCTGCGCAACTGACTCGGTACTACAAGATTGCACGCTCACTACA 598
|||||
192 ySTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIle 208
|||||
599 AGTGGCAGCTGGACCACTGTTTACAAAACAAATTTAGTCGAGTGATT 648
|||||
209 ileLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheG1 225
|||||
649 ATACTAGAAGATGATGATAAATGCTCCAGACTCTTTTGATTACTTTGA 698
|||||
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSers 242
|||||
699 GGCTCGACCTAGCTCTCATGATGAGGATAAAACCATATTGCTGCTTCAT 748
|||||
242 erTPsAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeu 258
|||||
749 CATGGAATGATAATGCAGAGACGAGTGTTCATGATGATCCCTATCGCTA 798

```

```
259  TyrArgSerAspPhePheProGlyLeuGlyTyrPheMetLeuSerLysSerTh 275
799  TACCGATCAGATTTTCTCGCTTGGTGGATGCTCAAGAGATCGAC 848
275  rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292
849  TTGGGATGAGTTATACCAAGATGGCCAAAGCCTTACTGGGATGATTGGC 898
292  euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal 308
899  TGAGACTAAAGGAAAAACCATAAAGCCGCCCAATTCATTCGACCGGAAGTC 948
309  CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePh 325
949  TGTAGAACATACAATTTTGGTGAACATGGGTCTAGTTTGGGACAGATTTT 998
325  eLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL 342
999  CAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGACGGTTGACTGGA 1048
342  ysSerMetAspLeuSerTyrLeuGluGluAspAsnTyrValLysHisPhe 358
1049  AAGCAAGGACCTGGGATACCTGCAGAGAGGGAACATACCAAGTACTTT 1098
359  GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLe 375
1099  TCTGCTTAGTCAGACAGACGACCAATTCAGGTTCTGACCTTGTCTT 1148
375  uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL 392
1149  AAAGGCTCAAAACATAAAGGATGATGTCGTATCCGGTATAAAGACCAAG 1198
392  euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGluTrpLys 408
1199  TAGAGTTTGAACCAATTCGACGGGAATTTGGTATATTTGAAGAATGGAAG 1248
409  AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheG1 425
1249  GATGCTGTGCTCGCAACAGCATATAAAGGAGTACTGTTTTCGAATCCA 1298
425  nTrpSerArgValPheLeuValSerProAspSerLeuArgGlnLeuG 442
1299  GACAACAAGCGTGTATTCCTGGTTGGGCCAGATTCGTAAATGCAGCTTG 1348
442  lyValGluAspThr 446
1349  GAATTCGAAATTCC 1362
seq_name: gb_pl:ATH249881
seq_documentation_block:
LOCUS   ATH249881
DEFINITION   Arabidopsis thaliana mRNA for N-acetylglucosaminyltransferase I
(GntI gene).
ACCESSION   AJ249881
VERSION     AJ249881.1 GI:18072832
KEYWORDS    GntI gene; N-acetylglucosaminyltransferase I.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
            1 (sites)
REFERENCE   Wenderoth,I., Tjaden,J. and von Schaewen,A.
            Isolation and characterization of different plant
            N-acetylglucosaminyltransferase I (GntI) cDNA sequences: Functional
            analyses in the Arabidopsis cgl mutant, and in potato and tobacco
            antisense plants
            Unpublished
JOURNAL    2 (bases 1 to 1830)
AUTHORS    von Schaewen,A.
TITLE      Direct Submission
JOURNAL    Submitted (28-SEP-1999) von Schaewen A., Pflanzenphysiologie,
```

```
Universitaet Osnabrueck, FB5, Barbarastr. 11, D-49076 Osnabrueck,
GERMANY
FEATURES             Location/Qualifiers
     source            1..1830
                        /organism="Arabidopsis thaliana"
                        /variety="Columbia"
                        /db_xref="taxon:3702"
                        /tissue_type="whole plant"
     gene              117..1451
                        /gene="GntI"
                        /EC_number="2.4.1.101"
                        /function="initiates complex-type glycans in the Golgi"
                        /codon_start=1
     CDS                117..1451
                        /product="N-acetylglucosaminyltransferase I"
                        /protein_id="CAC80700.1"
                        /db_xref="GI:18072833"
                        /translation="MARISCDLRFLLIPAAFMIYIQMLRFOTSOYADRLSSAIESE
                        NHCTSMRGLIDEVSIKOSRIVALEDMKNRODEELVOLKDLIQTFEKKGIKLTQSGQ
                        MPVAAVVMACSRADYLVRTVSVLYQTPVASKYPLFISQDSDQAVKSKSLYSNOL
                        TYMHLDPEFVPTERPEGLTAYTKIARHKYKWLQDLFYKHKFSRVIILEDDEMEIAPDF
                        FDYFEAAAALMDRDKTIMAASSNDNGQKQFVHDHPVALYRSDFPGLGMLKRTWDE
                        LSPKWPAYDDMLRLKHNKGRQFIAPVCRTYNFGEHSGSSLGQFFSOYLEPIKLND
                        VTVDWKAKDGLYLTGNTYTKYFSGLVQRAPITQGSDLVLKAQNIKDDRIKYKQDVEF
                        ERIAGEFGIFEWKQGVPRYAYKGVVVFRIQITRRVFLVGPDSVMQLGRNS"
BASE COUNT      540 a      348 c      422 g      520 t
ORIGIN
alignment_scores:
    Quality: 1768.00      Length: 438
    Ratio: 4.431          Gaps: 0
    Percent Similarity: 91.096      Percent Identity: 75.114
alignment_block:
US-09-591-466C-2 x ATH249881 ..
Align seg 1/1 to: ATH249881 from: 1 to: 1830
9  AspLeuArgTyrLeuLeuValValAlaAlaLeuAlaPheIleTyrIleG1 25
135  GACTTGAGATTTCTTCATCCCGGCAGCTTTTCATGTTCTACATCCA 184
25  nMetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAlaAla 42
185  GATCAGGCTTTTCAGACGCAATCACAGTATGCAGATCGCCTCAGTTCCG 234
42  laIleGluAlaGluAsnHisCysThrSerGlnThrArgLeuLeuIleAsp 58
235  CTATCGAATCTGAACCATTTGCACGTAGTCAAAATGCGAGGCCTCATAGAT 284
59  LysIleSerGlnGlnGlnGlyArgValValAlaIleGluGluGlnMetLy 75
285  GAAATTAGCATCAACACAGTCGCGGATTTGTCCTCGAAGATATGAAGAA 334
75  sHisGlnAspGlnGluCysArgGlnLeuArgAlaLeuValGlnAspLeuG 92
335  CCGCCAGGACGAAGACTTGTGCAGCTTAAGGATCTAATCCAGACGCTTG 384
92  luSerLysGlyIleLysLysLeuIleGlyAspValGlnMetProValAla 108
385  AAAAAAAGGAATAGCAAAACTCACTCAAGGTGGACAGATGCGCTGGGCT 434
109  AlaValValAlaMetAlaCysSerArgThrAspTyrLeuGluArgThrI1 125
435  GCTGTAGTGGTTATGGCCCTGCAGTCGTGCAGACTATCTTGAAGGACTGT 484
125  eLysSerIleLeuLysTyrGlnThrSerValAlaSerLysTyrProLeup 142
485  TAAATCAGTTTAAACATATCAAACTCCCGTTCCTTCAAAATATCCTCTAT 534
142  heIleSerGlnAspGlySerAsnProAspValArgLysLeuAlaLeuSer 158
```

```
|||||
535 TTATATCTCAGGATGGATCTGATCAAGCTGTCAAGACGACGATCAATTGAGC 584
|||||
159 TyrGlyGlnLeuThrTyrMetGlnHisLeuAspTyrGluProValHisTh 175
|||||
585 TATAATCAATTAACATATATGACGACCTTGGATTTTGAACGACGTTGTCAC 634
|||||
175 rGluArgProGlyGluLeuValAlaTyrTyrLysIleAlaArgHisTyrL 192
|||||
635 TGAAGGCCTGCTGACTGACTGCGTACTACAGATTGGCAGCTACATACA 684
|||||
192 ysTrpAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIle 208
|||||
685 AGTGGCAGCTGGACGAGTGTGTTTACAAACACAAATTTAGTCGAGTGATT 734
|||||
209 IleLeuGluAspAspMetGluIleAlaAlaAspPheAspTyrPheG1 225
|||||
735 ATACTAGAACGACGATATGGAAATGTCTCCAGACTTCTTTGATTACTTTGA 784
|||||
225 uAlaGlyAlaThrLeuLeuAspArgAspLysSerIleMetAlaIleSerS 242
|||||
785 GGCTGTCAGCTAGTCTCATGATAGGATATAAACCATATTATGCTGCTTCAT 834
|||||
242 erTrpAsnAspAsnGlyGlnArgGlnPheValGlnAspProAspAlaLeu 258
|||||
835 CATGGAATGATAATGGACAGAAGCAGTTTGTGCATGATCCCTATGCGCTA 884
|||||
259 TyrArgSerAspPhePheProGlyLeuGlyTyrPheMetLeuSerLysSerTh 275
|||||
885 TACCGATCACATTTTCTGCGCCTTGGTGGATGCTCAAGAGATCGAC 934
|||||
275 rTrpSerGluLeuSerProLysTrpProLysAlaTyrTrpAspAspTrpL 292
|||||
935 TTGGGATGAGTTATCACCAAGTGGCCAAAGGCTTACTGGGATGATTGGC 984
|||||
292 euArgLeuLysGluAsnHisArgGlyArgGlnPheIleArgProGluVal 308
|||||
985 TGAGACTAAGGAAACCATATAAGCGCGCAATTCATTGCACGGGAAGTC 1034
|||||
309 CysArgThrTyrAsnPheGlyGluHisGlySerSerLeuGlyGlnPhePh 325
|||||
1035 TCTAGAACATACAAATTTTGTGAACATGGTCTAGTTTGGGACAGTTT 1084
|||||
325 eLysGlnTyrLeuGluProIleLysLeuAsnAspValGlnValAspTrpL 342
|||||
1085 CAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTGACGGTTGACTGGA 1134
|||||
342 ysSerMetAspLeuSerTyrLeuLeuGluAspAsnTyrValLysHisPhe 358
|||||
1135 AAGCAAAGCACCTGGCATACCTGCAGAGGGAACATATACCAAGTACTTT 1184
|||||
359 GlyAspLeuValLysLysAlaLysProIleHisGlyAlaAspAlaValLe 375
|||||
1185 TCTGGCTTAGTCAGACAGACGACCAATTCAGGTTCTGACCTTGCTT 1234
|||||
375 uLysAlaPheAsnIleAspGlyAspValArgIleGlnTyrArgAspGlnL 392
|||||
1235 AAAGGCTCAAAACATAAAGGATGATGTCGTATCCGGTATAAAGACCAAG 1284
|||||
392 euAspPheGluAspIleAlaArgGlnPheGlyIlePheGluGlnTrpLys 408
|||||
1285 TAGAGTTTGACGCATTTGAGGGAATTTGGTATATTTTGAAGATGGAAG 1334
|||||
409 AspGlyValProArgAlaAlaTyrLysGlyIleValValPheArgPheG1 425
|||||
1335 GATGTTGTCACGACGACATATAAAGGATAGTAGTGGTTTTCGAATCCA 1384
|||||
425 nThrSerArgArgValPheLeuValSerProAspSerLeuArgGlnLeu 442
|||||
1385 GACAACAAGACGTGATTCTCTGGTTGGGCCACAGATTCGTATATGACGCTG 1434
|||||
442 lyValGluAspThr 446
|||||
```

1435 GAATTCGAATTC 1448

seq_name: gb_pat:A95057

seq_documentation_block:

LOCUS A95057 1854 bp DNA linear PAT 26-JAN-2000
DEFINITION Sequence 5 from Patent WO929879.
ACCESSION A95057
VERSION A95057.1 GI:6779211
KEYWORDS

SOURCE

ORGANISM

thale cress.

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 1854)

Von, S.A.

TITLE

VEGETABLE GNTI SEQUENCES AND THE USE THEREOF TO OBTAIN PLANTS WITH

A REDUCED OR LACK OF N-ACETYLGLUCOSAMINYLTRANSFERASE I (GNTI)

JOURNAL

Patient: WO 929879-A 5 17-JUN-1999;

VON SCHAEWEN ANTJE (DE)

FEATURES

source

1. 1854

/organism="Arabidopsis thaliana"

/strain="COLUMBIA"

/db_xref="taxon:3702"

/tissue_type="ALLE GEWEBE"

/clone_lib="LAMBDA UNI-ZAP (ECORI/XHOI) UND LAMBDA ACT"

/clone_lib="(XHOI)"

/clone_lib="(B) CLON(E): PBSK-ARA-GNTI-FULL #8"

/dev_stage="M: REIFE PFLANZEN"

19. .134

135. .1469

/gene="CGL"

/gene="CGL"

/standard_name="GNTI"

/EC_number="2.4.1.101"

/function="INITIIERT KOMPLEXE N-GLYKANE AUF SEKRET."

GLYKOPROTEINEN"

/note="ERSTE GNTI-SEQUENZ AUS ARABIDOPSIS (UNPUBLIZIERT)"

/codon_start=1

/evidence=experimental

/product="BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE I"

/protein_id="CAB70466.1"

/db_xref="GI:6779212"

/translation="MARISCDRLFLIPAAFMFIYQMRLFTQSOYADRLSSAIESE

NHCTSQMRGLIDEVSIKQSRIVALEDMKRDEELVOLKDLIOTFEKKGAKLTQGGQ

MPVAAYVVMACSRADYLERTVKSVLTVPVASKYPLFI SDGSDQAVKSKSYNQL

TYMHLDPEPVTERPCELTAYYKIARHYKWLQDLFYKHKFSRVILEDDMETADPF

FDYFEAAASLMDRDKTIMAASSWNGQKQFVHDHPYALYRSDFPGLGMLKRTWDE

LSPWPKAYWDDMLRLKENHKGFIAPVCRTYNFEHGSSIGQFFSQYLEPIKLNLD

VTVDKAKGLGLEYNTKYFSGLVQRQARPIQGSDLVLKAQNIKDDIRIRYKQVEF

ERIAEGFGEEMKDGVPRTAYKGVVVFRIQTTRRVFLVGPDSVMQIGIRNS"

<157. .>215

/gene="CGL"

/standard_name="MEMBRANKER EINES TYP II GOLGI-PROTEINS"

/function="MEMBRANKER (AMINO ACID 8-27)"

/note="DURCH VERGLEICH MIT TIERISCHEN GNTI-SEQUENZEN

IDENTIFIZIERT"

/codon_start=2

/product="HYDROPHOBER AMINO ACIDBEREICH IN GNTI"

/protein_id="CAB70467.1"

/db_xref="GI:6779213"

/translation="EISSHPGSGFHVHLHPDEAF"

1470. .1848

545 a 357 c 430 g 522 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 1768.00

Length: 438

FEATURES	source	Location/Qualifiers
		1. .147215
		/organism="Arabidopsis thaliana"
		/variety="Columbia"
		/db_xref="taxon:3702"
		/chromosome="4"
gene		complement(join(830. .2566,2624. .3049))
gene		/gene="F20D10.10"
		830. .3049
		/gene="F20D10.10"
CDS		complement(join(830. .2566,2624. .3049))
		/gene="F20D10.10"
		/note="strong similarity to retrotransposon -like protein
		-Arabidopsis thaliana, PID:e1248853"
		/codon_start=1
		/product="putative protein"
		/protein_id="CAB37529.1"
		/db_xref="GI:4467095"
		/translation="MMNGLRRTFWSSIHKKKDNNRVDDSLDRQKPTTTSRRGFFSNPS TPRSETPDSVTCSPITPCQWSATAISTPSPSLPASKLQCDTSGDVTPTNRSPLS FLVSSSSSTPSPKSPASFLLSKLCTFKVRIINRNQSSISSKGCICLQSAKAG RGTAIFPACSHTFHFPCVASRAGRNLLSDCPVCASWRETSLPLSLSSLHESGS ESDKITRESKNNKSLRVYNDDEPLISSPISRTGNTIPESNEDEEEDNDGEGKGF YNTFSLPTKMLTDSVTGHVDKLVSEAAIVAVGRNETYSVLWKIKSPSLPTARR SPDLVTVIDVSGGNIEMVKRAMROVISLRETDRLSMVSFSSSKRLTPLRMTANG RLRLARRIVDDISGDDGMSVNDVAKKAVIEDRQKLFITFVLTDRNRNSAHQAO LAQDPVTFSTRSHLEIPHTITWLGCHNALPEDVFAKRIKLSLVODUTNLNLGLV SGSGGQKTVSVLSGRPVWLSGLIRGDMYGDDEEREVLVELKSPSSRSQRIIMTV SHRVDPTQEIKNYGDRLMTPPTTVTSSDPSIARLNLHVSTRVAESRRLVEVND YSGAERMARALLVQGLSSDCLRGLEALADLNLRLGRHVAVSPPEVPQKSE PLTPTSAMRAERLAKVAIMRKMNRVSDLHGFENARF"
exon		complement(830. .2566)
		/gene="F20D10.10"
		/number=1
intron		complement(2567. .2623)
		/gene="F20D10.10"
		/number=1
exon		complement(2624. .3049)
		/gene="F20D10.10"
		/number=2
repeat_region		5063. .5098
exon		/note="35bp ta tandem repeat"
		complement(9755. .10264)
		/gene="F20D10.20"
		/number=1
gene		9755. .12463
		/gene="F20D10.20"
gene		complement(join(9755. .10264,10338. .11657,11741. .11989, 12071. .12191,12300. .12463))
		/gene="F20D10.20"
CDS		complement(join(9755. .10264,10338. .11657,11741. .11989, 12071. .12191,12300. .12463))
		/gene="F20D10.20"
		/note="similarity to Caenorhabditis elegans cosmid F32B5, gene F32B5.7, PID:g2088783
		Contains Eukaryotic putative RNA-binding region RNP-1 signature [KGSCLPDM]
		contains EST gb:N97112:AA728688"
		/codon_start=1
		/product="putative protein"
		/protein_id="CAB37530.1"
		/db_xref="GI:4467096"
		/translation="MKEKEQTLEWNEAQKIDISVLLAAAKKHLFLGAVDRNRCILY DGPALQRAIYRYNAILPLQAQTESSICQGLVPPLDCEWYWHCHRLNPVRYKTDG EQFYDLVDNGSVSSVNGKQSQTETLMKRLYPTPEYDLDFANAISEPADVSALSKC TYDVLVLRQSPFYQVSRSHVDNDVLFQEAARYKAFYLIKGNRRERSIKLFCVP TYDIDLTHWHQIHAISYCNLDLTKMIGKVLEHDDTSDRSKGLDTPGSGTAQWEE TFGRYVWAKAMKNGNTPKVTTPVYSPVCKKSKIAKEESONVIOQPEVKYFIEVLEI VGVKNLPDAHKGVFVLEFSKTPQDSLFPNERRLTVLSESGEKQVAFQCEPTGELSF QLMSSKSLGFTLSFSBELSPVTKLSVEKMLELPTTKRGKADDPNPISLRVAVSFT PPTRSPTVLHVQARPSLQSGCPLPMLKRSFTRVYDETEFVILNQRNSNDA APGRDRQVIGVKECGEYTLAEYDGTFWSLDLSKSLQTCNPTADGPLFELSGTRM KVYYSRKLKEYPKHCKSLRSEQDFMTAVEFSKQHPYKAVGLGLDLKFGSIEANEKWL

VLPGVMSFILLSLLKKEGFSAAKDTYKANGITEESTIDVLSQEKLEETNMMDVDT
TTPVAVAAEKINGGARCFSKELSGNMIEEGGHCAGCGGCGGGGGGGGRCGGM
TKLEGCGSGCTGGSTGCNCGCGGCMKNNANGNAPSVDNAVTA"

complement(10265. .10337)

/gene="F20D10.20"

/number=1

complement(10338. .11657)

/gene="F20D10.20"

/number=2

complement(11658. .11740)

/gene="F20D10.20"

/number=2

complement(11741. .11989)

/gene="F20D10.20"

/number=3

complement(11990. .12070)

/gene="F20D10.20"

/number=3

complement(12071. .12191)

/gene="F20D10.20"

/number=4

complement(12192. .12299)

/gene="F20D10.20"

/number=4

complement(12300. .12463)

/gene="F20D10.20"

/number=5

complement(join(13386. .14030,14118. .14738,14828. .15394,
15491. .15569,15669. .15730,16043. .16069))

/gene="F20D10.30"

13386. .16069

/gene="F20D10.30"

complement(join(13386. .14030,14118. .14738,14828. .15394,
15491. .15569,15669. .15730,16043. .16069))

/note="strong similarity to dnaK-type molecular chaperone
PHSP1 -Pisum sativum, PID:g20835

Contains ATP/GTP-binding site motif A (P-loop) [AEAYLGKS],
Heat shock hsp70 proteins family signatures [IDLGTNS],
Hsp70_2 [VFDLGGTFDVSI], Hsp70_3 [VLLGGMTRVPKVOE],
contains EST gb:T4452, Z24523, T04782, Z17893, Z17973"

/codon_start=1

/product="heat shock protein 70 like protein"

/protein_id="CAB37531.1"

/db_xref="GI:4467097"

/translation="MASVSAPKSVSANGKSMFCKLGLARPCSPVGNVDVICIDLG
TTNSCVSMGKTARVIEAEGSRTPSVVAMKQKCELLVGTFAKRAVNTNPTIFG
SKRLIGRRFDDPOTQEMKWPYKIVKAPNGDAWANGKFSQIGANVLTKMRKT
AEAYLGKSINKAVYVPAYFNDQAKQKADGATAGLDVQRIINEPTAAALSVMNNK
EGVIAVPDLGGTDFDVSILEISSGVFEVKATNGDTFLGGEDFNTLLEYLVNFKKSD
NIDLTDLNALQRLREAAEKATIELSTTQTEINLPTITADASCAKHLNLTLSKPKPE
GLVGLKLIERTSPQCNCIDAGVTIKEDVLLVGGTRVPKVOEIVSEIFGKSPCKG
VNPDEAVANGAAIQGGILRGVDKDLLLDVPLSLGLTGLAVETKLIIPNTTIPPKK
SQVSTEAANDNMQVGIQLQGERMAADNKLGEFOLVGIIPAPRGMPOLEVTIDDA
NGITTVSKADKATGKEQNTIRSSGSLSDDEINRMVKEALNAKQDEKQLIDLKNS
ADPTIYVSKSLSEYREKIPAEIATSEIETAVSLDRTAMAGEDVEDIKAKVEAANKAVS
KIEHMSKSGSGSGSGSGTGTGTEQTPAEFEFEASGSRK"

complement(13386. .14030)

/gene="F20D10.30"

/number=1

complement(14031. .14117)

/gene="F20D10.30"

/number=1

complement(14118. .14738)

/gene="F20D10.30"

/number=2

complement(14739. .14827)

/gene="F20D10.30"

/number=2

complement(14828. .15394)

/gene="F20D10.30"

/number=3

complement(15395. .15490)


```
/gene="F20D10.30"
/number=3
complement(15491..15569)
/gene="F20D10.30"
/number=4
complement(15570..15668)
/gene="F20D10.30"
/number=4
complement(15669..15730)
/gene="F20D10.30"
/number=5
complement(15731..16042)
/gene="F20D10.30"
/number=5
complement(16043..16069)
/gene="F20D10.30"
/number=6
16556..19503
/gene="F20D10.40"
complement(16556..16790)
/gene="F20D10.40"
/number=1
complement(join(16556..16790,16869..17056,17133..17377,
17453..17618,17710..17975,18152..18358,18789..19503))
/gene="F20D10.40"
complement(join(16556..16790,16869..17056,17133..17377,
17453..17618,17710..17975,18152..18358,18789..19503))
```

alignment_scores:

```
Quality: 1133.50      Length: 1065
Ratio: 2.806          Gaps: 27
Percent Similarity: 37.934  Percent Identity: 31.174
```

alignment_block:

US-09-591-466C-2 x ATF20D10/rev

Align seg 1/1 to reverse of: ATF20D10 from: 1 to: 147215

```
9 AspLeuArgTyrLeuLeuValValAlaLeuAlaPheIleTyrIle.. 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123209 GACTTGAGATTCTCTCATCCGGCAGCTTTCATGTCATCATCA 123160
24 .....
123159 GGTCAATCTATATCCGATGATCGCTTTTGAAGTTGAAACTGAAATG 123110
24 .....
123109 ATTCATCGCTGCAACCTAATCTGCTGTGTAATATCCTCTCTCTTC 123060
25 Gln..MetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAla 40
::: |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
123059 TCTAGATGAGGCTTTTCCAGACGAATACAGTATGCAGATCGCCTCAGT 123010
41 AlaAla.....
::|
123009 TCCGCTGTGAGTTCTACITTTTACCACCTGCCTTTTGTCTTTAGACCATG 122960
42 .....
122959 CTACTAGTTCCTTTTGAATCCCATTTTGGCTCTTACAGCCCATGA 122910
42 .....
122909 ACTACTTTCAAGAAGACACTATCTGCACATTACACCATATCATTCG 122860
42 .....
122859 CTTTGCTATTCGATTTTCTCTGGTGTGTGATGAAGATCTATACACAGC 122810
43 .....
122809 CTTTCATTTGTTCTCCAGATGAATCTCAGAACCATTCGACTAGTCAAA 122760
```

```
53 hrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValValAla 69
||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||
122759 TCCGAGGCTCTATAGATGAAGTTAGCATCAAAACAGTCGCGATTGTGCC 122710
70 LeuGlu..Glu.....
|||||
122709 CTCGAAGGTATATCCCTTCTCTTCTCATTTGGGACTTGTGGAACTCT 122660
72 .....
122659 AGATGACCATGGTTCTTGGAACTTATTGAATACATAAAAAATTAACTTT 122610
73 .....GlnM 74
122609 CTAATGGAGTAGAAGACTTATACATTGATTTCTGAGTGTCTGTAGATA 122560
74 etLysHis...GlnAspGlnGluCysArgGlnLeuArgAlaLeuValGln 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||
122559 TGAAGAACCGCCAGGAGAACTTGTGCAGCTTAAGGATCTAATCCAG 122510
90 AspLeuGluSer.....
:::|||||
122509 ACGTTTGAAGTGCCTTACTATCATCCTGTAAACTCATTTTCTGCTTTC 122460
94 .....Lys.Gly 95
||||| ||| |||
122459 TTTTATTAGTTTCTACTAATGCTTATTCTTCTCTCCAGAAAGGA 122410
96 IleLysLysLeuIle.....GlyAspVal.....
||| ||||| |||:||||
122409 ATAGAAACTCTCACTCAAGGTGGACAGTCTGGATAAGAAGATCGCTTTC 122360
103 .....
122359 TTCGATTGTGAACCTGCTATCACTTGTCTATATATATACATCTAACCTGTT 122310
104 .....GlnMetProValAlaAlaValValValMetAla 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||
122309 TCTGTGTTTTTATTAGCAGATGCCTGTGCTGCTGTAGTGTATGGCC 122260
115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu..... 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||
122259 TGCAGTCGTGCACACTATCTTGAAGAGACTGTAAATACAGTTTAAAGTG 122210
129 .....
122209 AGATACTATACTGATTAACTTATACCACATTTTGTGATTGAACATCTA 122160
130 .....Lys..TyrGlnThrSerValAlaSerLysT 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||
122159 ATGGCTATTTCTTTTATACAGATATCAAACTCCCGTTGTCTCAAAAT 122110
139 yrProLeuPheIle.....
|||||:|||||:|||||:|||||:|||||:|||||:|||||
122109 ATCCTCTATTTATATCTCAGGTACTTAATTTCCAGGCTGGTGGTCTCTTA 122060
143 .....
122059 GACTGTTTTATATCAACGAATTACTAACCATGGTCTTTTCCAACGACAT 122010
144 ....Ser..GlnAspGlySerAsnProAspValArgLysLeuAlaLeuSe 158
||| |||||||:|||||:|||||:|||||:|||||:|||||:|||||
122009 ATTTTCTGACAGGATGGATCTGATCAAGCTGTCAAGAGCAGCATCTAG 121960
158 rTyrGlyGlnLeuThrTyrMet.Gln.....
|||||:|||||:|||||:|||||:|||||:|||||:|||||
121959 CTATAATCAATTAACATATATGTCAGGTAATTTAATCTAGCGTGATTTTC 121910
166 .....
121909 TCCTCACTGTTCAAGTAGATAAAATGAGGATGAAGGCTCGTTTTCCCTTT 121860
```

```

167 ....HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182
121859 TCAGCACTTGGATTGGAACCACTGCTCACTGAAGAGCCTGGCGAACTGA 121810
182 aAlaTyrTyrLysIleAlaArg..... 189
121809 CTGCGTACTACAAGATTGCAGTAAGGACTGGTTCTCTTTCTCTGCGT 121760
190 .....HisTyrLysTr 193
121759 CCCTCAGCTCCATTACTGAACACTTTTAAATACAGGTCACACTACAAGTG 121710
193 pAlaLeuAspGlnLeuPheHisHisAsnPheSerArgValIleIleL 210
121709 GGCACCTGGACCAGTTCTTTTACAAACACAAATTTACTCGAGTGATTATAC 121660
210 euGlu..... 211
121659 TAGAAGGTGGCGCTTTTAAACCTTTTCACACTCATCGTACCTTGGCTGTGT 121610
211 ..... 211
121609 ATTGTTTAAGAGGGCTGTCTTTTACCTTGTGGAGTATTGCTTAAAGAGG 121560
212 .....AspAspMetGluIleAlaAlaAspPhePheA 222
121559 TTATCTTGCTCTTTTAAACAGATGATATGAAATTTGCTCCAGACTTCTTTG 121510
222 sPyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235
121509 ATTACTTTGAGCTGCACCTAGTCTCATGGATAGGATAAGTATGTTACC 121460
235 ..... 235
121459 TACTGAAGCTAGACTCTTTATAGTATAAGTTTCCACAGAGTATACAAATGT 121410
236 .....SerI 237
121409 TTTTATTACTTTTCTCATCTCTTTCCTGGTCCAAACATGTGTTAGAACC 121360
237 leMetAlaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253
121359 TTATGGCTGCTTTCATCATGGAATGATAATGGACAGAAGCAGTTGTGCAT 121310
254 AspPro..... 255
121309 GATCCCTGTAGATTCTTCTTCTTCTTCTTTTACCTTGTCTTCTCTCT 121260
255 ..... 255
121259 TGCGTGTCTTTCAATACCTCATTATGCTCTAGCTCATGCCAGCTAAATA 121210
256 .....Asp..AlaLeuTyrArgSerAspPhePheProGlyLeuGlyTrpM 270
121209 ATCTGCAGATGCCCTATACCGATACAGATTTTTCCTGGCTTGGGTGGA 121160
270 etLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLys... 285
121159 TGCTCAAGAGATCCACTTGGGATGAGTTATCCACCAAGTGGCCAAAGGCA 121110
285 ..... 285
121109 TATCCTTATATGCTTTTCTTTTCTTCAATCTCGACTATAGTTTGAGC 121060
285 ..... 285
121059 CTCTTTAAATCTTAAAAAATAAGTTTGACGTTCTTAAGACTTTTCT 121010
286 .....Ala..TyrTrpAspTrpLeuArgLeuLysGluAsnHi 298
121009 TAACAAATATCCACTTACTGGGATGATTGGCTGAGACTTAAAGGAAACCA 120960
298 sArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheG 315

```

```

120959 TAAAGCCGCCCAATTCATTGCGACCGGAAGTCTGTAGAACATACAATTTG 120910
315 lyGluHis..... 317
120909 GTGAACATGTTAGTACCCGATTATGCTATCCTATATATCATAGTTTACA 120860
317 ..... 317
120859 AGTCGTGTTTTTGTAAATGCTGTTAAATGACCTATGTTTCGTCGATATGAC 120810
317 ..... 317
120809 TTCTGCTTTCTGTATGAAAAATGATTATTATGACGCTTCGTCCTATAGTTT 120760
317 ..... 317
120759 CCCGACACATGGATGTAGACAGTCAATACCACATAATATATTTTGGCATT 120710
318 .....GlySerSerLe 321
120709 TCATATTATCCATAACCTTGTATATTAATTTGGCAGGGGTCT.AGTTT 120661
321 uGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspVal. 337
120660 GGSACAGTTTTTTCAGTCAGTATCTGGAACCTATAAAGCTAAACCATGTGA 120611
337 ..... 337
120610 CGGTAAGCGTAAGTGTGTTTTTTTACCCCTTTTGTGTCATCAGTATGTTT 120561
338 .....GlnValAspTrpLysSerMetAspLe 346
120560 CACTGGTTTCTCTATTTTCATTATAAAGTTGACTGGAAAGCAAAGGACCT 120511
346 uSerTyrLeuLeuGlu..... 351
120510 GGGATACCTCAGAGAGGTAACATTTACAGAACTCCGGCCCTTTTTTTTA 120461
351 ..... 351
120460 TACCATTTTGCATGTTTTTGAGTGGAGAAATGTAGTTTTTAACTGTCTC 120411
352 .....AspAsnTyrValLysHisPheGlyAspLeuVally 363
120410 TCATGTCTATCAAGGGAACCTATACCAAGTACTTTTCTGGCTTAGTGAG 120361
363 sLysAlaLysProIleHisGlyAlaAlaAspAlaValLeuLysAlaPheAsnI 380
120360 ACAAGCAGCAACCAATTCAAGTTCTGACCTTGTCTTAAAGGCTCAAAACA 120311
380 leAspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAsp 396
120310 TAAAGATGATGTTGCTATCCGCTATAAAGACCAAGTAGAGTTTGAACGC 120261
397 lleAlaArgGlnPheGlyIlePheGluGlnTrpLys..... 408
120260 ATTCGAGGGGAATTTGGTATATTTGAAGNATGGAAGGTAATTTGAGATG 120211
408 ..... 408
120210 ATTGAATATAATCTTCAATAATTGTAGTCCAAATGGGTCACTTTTTTT 120161
409 .....AspGlyValProArgAlaAlaTyrLysGly 418
120160 GTACCCCTTTGATTGTGTAGGATGTTGCTCGAACAGCATATAAAGGA 120111
419 lleValValPheArgPheGlnThrSerArgArgValPheLeuValSerPr 435
120110 GTAGTGTGTTTCCGAATCCAGACAAACAGAGCTGTATTCCTGGTTGGGCC 120061
435 oAspSerLeuArgGlnLeuGlyValGluAspThr 446
|||||:  |||||:

```

```

120060 AGATTCTGTAATCCAGCTTGGAATTGGAATTCC 120027
seq_name: gb_pl:ATCHRIV89

seq_documentation_block:
LOCUS ATCHRIV89 199789 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 89.
ACCESSION AL161593
VERSION AL161593.2 GI:7270796
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 199789)
Murphy, G., Ridley, P., Hudson, S., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
2 (bases 195922 to 196568)
Robben, J., Braeken, M., Gymonprez, B., Volckaert, G., Mewes, H.W.,
Lemcke, K., and Mayer, K.F.X.
Unpublished
3 (bases 1 to 68137; 40958 to 134717; 99060 to 190997)
Wedler, H., Kutzner, M., Wambutt, R., Mewes, H.W., Lemcke, K. and
Mayer, K.F.X.
Unpublished
4 (bases 1 to 199789)
EU Arabidopsis sequencing, project.
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitze 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV88 at the 5' end and an
overlap with ATCHRIV90 at the 3' end.
FEATURES
source
Location/Qualifiers
1..199789
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(5433..5537)
/gene="AT4g38130"
/number=1
5433..7997
/gene="AT4g38130"
complement(join(5433..5537,5612..5683,6112..6169,
6258..6328,6584..6814,6945..7385,7470..7997))
/gene="AT4g38130"
complement(join(5433..5537,5612..5683,6112..6169,
6258..6328,6584..6814,6945..7385,7470..7997))
/gene="AT4g38130"
/number=1
/contains EST qb:R90034, AI994124.1, W43397,
AW004334.1, F15302, AA605428"
/codon_start=1
/product="Histone deacetylase"
/protein_id="CAB80478.1"
/db_xref="GI:7270797"
/translation="MDTGGNSLASGPDGKVRKVCYFYDPEVGNVYVYCGHPMKPHRIR
MTHALLHYGLQHMVLKPPARDRLCRFHADYVFLRSITPEQQOIRKOLKRF
NVGDCPVFDGLYSCOTYAGSGVSKNLHGLCDIAINWAGLHHAQKCEAGFCY
VNDIVLAILLELLOKHVERLVYDIDHGGDGEAFATYATDRMTVSPFKFGDYFGTGH
IQDIGSGGATYSLNVPDLDGIDDESYYHLLEKFKPMGKVFIFRPGAVVLCQGDLSLG
RMCPHEYYEYFGPDYTLHVPASNNKSNRMLLEIRNLDLHNLKLOHAPSVPFOER
PPDPTPEVEDDEQDGDKRWDPDSMDVDDDRKP IPSRVKREAVEPDTKDKGLKIM
ERKGCGEVEVDESGSTKVTGVPNVPVVEASVKWEEECTNGGAEQAFPPKT"

```

DAVATCSBLESHGSPNPVPTFVELVDALCRVKVGEQAQAIDTLNOKGFAVNVKAVKE
 FNDKRAPPSPSLAWAEIFKKKPTKPF"
 11352..12353
 /gene="AT4g38160"
 exon 11352..12353
 /gene="AT4g38160"
 /number=1
 CDS 11352..12353
 /gene="AT4g38160"
 /note="contains EST gb:AI992903.1, H37273, A1100213, T41873, AA067478, A1100538"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB80481.1"
 /db_xref="GI:7270800"
 /translation="MEVNTSSIMWFFRDKGDDPSIDKMLRKCKQLEKAQSDVASEN WYLSNIVGIQERKLPYIVSRCPKILTLRLERLIPMEVCESSLRGNPREVASAITKF PPILSHVEEKLLPILAFQALGVPEPTOLGRMLFNPRLLSYSDITKLTVIVSFLASL GLDQGMIGKVLVNPFLMGYSVDKRLRPTEFLKSSVGLSEDKISVMVMPOLLCL DVNKLKPNYDYLKECGGDSQIATMTVGPQIILKSVKNSLQPRIRFLVQVMGRGMD EVASYPEFFHHGLKKKVESREKLVKKNNIDCSLREMLDCNTKKFHEKFGFSEVTASF" 13703..15379
 /gene="AT4g38170"
 exon 13703..15379
 /gene="AT4g38170"
 /number=1
 CDS join(13703..15130,15212..15379)
 /gene="AT4g38170"
 /note="Contains Prokaryotic membrane lipoprotein lipid attachment site AA73-83 contains EST gb:N37149, T14014"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="CAB80482.1"
 /db_xref="GI:7270801"
 /translation="MSRVHVLNLYKRRLENPGFLVAIEDDCGNVFWADPTCRLNYT YFGDTLVFDITRRGKRYQVFPFAFTGNHHGQPVLEGCALLNESSESFANLQTWL QAMAPPPPSITVPEPDLIOVANSRVFSQRLRFSPQLIFEETEEKLAHVFOAHPTE SEFTNCUTEETAEPEASWDSIVRRYIMEDNDWLQSIYNARQOMVRVFIIRDTPYGL STNGSSILNSFFGFDVASTTMOMLIKOYEKAIDSWREKELKADYEATNSTPYMKTP SPMEKQASLYTAAFTFKQEEFVETLAI PANIISDSGTHHTYRVAKFGEVHKGHTVS FDSLEVKANCSOMFESGILICRHILAVFSAKNVLLPSRYLLRRWTKKAKIRCTEQ PFNSNGQESLNLFCFNSLRQETAKYVEEGAKSIQIYKVAMDALDEAAKKVAAASNRP GTRLNPGEAIVPSEAREATANATHPGGEKERTILELTAELERTQRCVEVYRANLLSIL RDMEEQFQLSLKQVARNLSIKE" 15131..15211
 /gene="AT4g38170"
 /number=1
 exon 15212..15379
 /gene="AT4g38170"
 /number=2
 exon complement(15642..15815)
 /gene="AT4g38180"
 /number=1
 gene 15642..18344
 /gene="AT4g38180"
 gene complement(join(15642..15815,16152..18344))
 /gene="AT4g38180"
 CDS complement(join(15642..15815,16152..18344))

alignment_scores:

Quality: 1133.50 Length: 1065
 Ratio: 2.806 Gaps: 27
 Percent Similarity: 37.934 Percent Identity: 31.174

alignment_block:

US-09-591-466C-2 x ATCHRIV89/rev ..

Align seg 1/1 to reverse of: ATCHRIV89 from: 1 to: 199789

9 AspLeuArgTyrLeuValValAlaAlaAlaPheIleTyrIle... 24

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

44131 GACTTGACATTCTCTCATCCCGCAGCTTCATGTTCTATCATCCA 44082

24 24
 44081 GGTCAATCTATATCCGATGATCTCGCTTTTGAAGTTGAAACTGAAATTTG 44032
 24 24
 44031 ATTCCATCGTGCACCACTAAATCTGTCTGTAATTATCTCTCTCTCTTC 43982
 25 Gln..MetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAla 40
 ::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43981 TCTAGATGAGGCTTTTCCAGACGCAATCACAGTATGCAGATCGCCTCAGT 43932
 41 AlaAla..... 42
 :::|||
 43931 TCCGCTGTGAGTTTCTACTATTTTACCACCTGCCTTTTGTCTTTAGACCATTTG 43882
 42 42
 43881 CTACTAGTCTCTTTTGAATCCCATTTTGGCTCTTTTACAAGCCCATGA 43832
 42 42
 43831 ACTACTTTCAAGAAGAAGACACTATCTGCACATATACACCATATCATTCG 43782
 42 42
 43781 CTTTGTGATTTTCGATTTTTCCTGGTGTGTGATGAAGATCTATACACAGC 43732
 43IleGluAlaGluAsnHisCysThrSerGlnT 53
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43731 CTTTCATTTGTTCTCCAGATCGAATCTGAGAACCATTCGCACTAGTCAAA 43682
 53 hrArgLeuLeuIleAspLysIleSerGlnGlnGlnGlyArgValValAla 69
 ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43681 TCGAGGCCTCATAGATGAAGTTAGCATCAACACAGTCGCGGATTTGTTCG 43632
 70 LeuGlu..Glu..... 72
 |||||
 43631 CTCGAAGGTATATCCCTCTCTCTTCTTCATTCGGCATTGTGGAATCCT 43582
 72 72
 43581 AGATGACCATGTTCTTGGAACTATTGAATCTATACAAAATTAACTTT 43532
 73Gln 74
 :::|
 43531 CTAATGAGTAGAGACTTATACATTGATTCTGAGTGTCTGTGTAGATA 43482
 74 etLysHis...GlnAspGlnGluCysArgGlnLeuArgAlaLeuValGln 89
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43481 TGAAGAACCAGCCAGGACGAGAACTTGTGCAGCTTAAGGATCTAATCCAG 43432
 90 AspLeuGluSer..... 93
 :::|||
 43431 ACGTTTGAAGTGCCTTACTATCACCCTGTAACACTATTTCTTCTGCTTTC 43382
 94Lys.Gly 95
 |||||
 43381 TTTTATTAGTTTCTACTAATGCTTATTTTCTCTCCAGAAAAAGGA 43332
 96 IleLysLysLeuIle.....GlyAspVal..... 103
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43331 ATAGCAAAACTCCTCAAGGTGGACAGGCTCTGGATAAGAAGATGCTTTC 43282
 103 103
 43281 TTCGTATTGTAACTGCTATCATTGCTATATATATATATATATCTAACCTGTT 43232
 104GlnMetProValAlaAlaValValMetAla 114
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 43231 TCTTGTGTTTATTATAGCAGATGCCTGTGGCTGTGTAGTGTGTTATGCC 43182

115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu..... 129
|||||
43181 TGCAGTCGCGAGACTATCTTTGAAAGGACGTGTTAAATCAGTTTTTAACGTG 43132
129 129
43131 AGATACTATACTGATTAAATCTTTATACCACATTTTGATTCACATATCTA 43082
130Lys...TyrGlnThrSerValAlaSerLysT 139
43081 ATGCTATTCTTTTTTATACAGATATCAAACTCCCGTTGCTTCAAAAT 43032
139 yrProLeuPheIle..... 143
43031 ATCTCTATTATATCTCAGGTACTTAATTTGCAGGCGTGTCTCTTA 42982
143 143
42981 GACTCTTTTATATTCAACGAATTACTAACCATGGTCTTTTCCAAAGGACAT 42932
144Ser...GlnAspGlySerAsnProAspValArgLysLeuAlaLeuSe 158
|||||
42931 ATTTTTCAGAGGATGATCTGATCAAGCTGCTCAAGAGCAAGTCAATGAG 42882
158 rTyrGlyGlnLeuThrTyrMet..Gln..... 166
|||||
42881 CTATAATCAATTAACATATATGACGTAATTTAATCTAGCGTGATTTTC 42832
166 166
42831 TCCTCAGTCTCAAGTAGATAAATGAGGATGAAGGCTCGTTTTCCCGCTTT 42782
167HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182
|||||
42781 TCAGCACTTGGATTTTGAACCACTGGTCTCACTGAAGGCCCTGGCGAACTGA 42732
182 alaIaTyrTyrLysIleAlaArg..... 189
:|||||
42731 CTGCGTACTACAAGATGTCACGTGAAGGACTGGTTCTCTCTCTGCGT 42682
190HistyrLysTr 193
42681 CCCTCCAGCTCCATTACTGAACACATTTTAAATACAGGTCACACAAAGTG 42632
193 pAlaLeuAspGlnLeuPheHisLysHisAsnPheSerArgValIleIleL 210
|||||
42631 GGCACCTGGACCACTTGTTTTACAAACACAAATTTAGTCGAGTGATTATAC 42582
210 euGlu..... 211
42581 TAGAAGGTGCGGCTTTTAACTTTTCACACTCATCTGCTGCTGTTGT 42532
211 211
42531 ATTGTTTAAGAGGGCTGTCTTACCTTGTGAGTATTGCTTAAAGAGAG 42482
212AspAspMetGluIleAlaAlaAspPheA 222
42481 TTATCTGTCTCTTAAACAGATGATGGAATTTGCTCCAGACTCTCTTTG 42432
222 spTyrPheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235
42431 ATTACTTTGAGGCTGACGCTAGCTCTCATGGATAGGATAAGTATGTTACC 42382
235 235
42381 TACTGAAGCTAGACTCTTATAGATAAGTTTCCACAGAGTATACAAATGT 42332
236SerI 237
42331 TTTTATTACTGTTTCTCATTTCTTTCTGGTCCAAACATGTGTAGAACA 42282
237 leMetaIaIleSerSerTrpAsnAspAsnGlyGlnArgGlnPheValGln 253

42281 TTTATGGCTGCTTCATCATGGAATGATAATGACAGACAGTGTGTGCAT 42232
254 AspPro..... 255
42231 GATCCCTGTAAAGATTCTTCTTCTTCTTCTTACCTTGCCTCTCTCT 42182
255 255
42181 TGCTGTCTTCTTCAATACCTCATTATGCTAGCTCATCGCAGCTAAAATA 42132
256Asp...AlaLeuTyrArgSerAspPhePheProGlyLeuGlyTrpM 270
42131 ATCTGAGATGCGCTATACCGATCAGATTTTTCCTGGCCTTGGGTGA 42082
270 etLeuSerLysSerThrTrpSerGluLeuSerProLysTrpProLys... 285
42081 TGTCAAGAGATCGACTTGGGATGAGTTATCACCAAAAGTGGCCAAAGGCA 42032
285 285
42031 TATCCTTATATGCTTTTTTCTTCTTCAATCTCGATGACTATAGTTTGACG 41982
285 285
41981 CTTCTTAAATCTTAAAAAATAAGTTTGTGAGCTTTCTTAAGACTTTTCT 41932
286Ala..TyrTrpAspTrpLeuArgLeuLysGluAsnHi 298
41931 TAACAAATATCCACTTACTGGGATGATTGGCTGAGACTAAAGCAAAACCA 41882
298 sArgGlyArgGlnPheIleArgProGluValCysArgThrTyrAsnPheG 315
41881 TAAAGCCGCCAATTCATTCGACCGGAAGTGTGTAGAACATACAATTTTG 41832
315 lyGluHis..... 317
41831 GTGAACATGTTAGTACCCGATTATGCTATCTATATATATCATAGTTTACA 41782
317 317
41781 AGTCTGTTTTTGTAAATGCTGTTAAATTTGACCTATGTTTCGTCGATGAC 41732
317 317
41731 TTCTGCTTCTGTATGAAAAATGATTATTGCGACTTCGCTCCTATAGTTT 41682
317 317
41681 CCGACACATGGATGTAGACGTCAATACCACATAATATATATTTTGCATT 41632
318GlySerSerLe 321
41631 TCATACTTATCCATAACCTTGTATTATTAATTTTGGCAGGGGTCT..AGTTT 41583
321 uGlyGlnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAspVal. 337
41582 GGGACAGTTTTTTCAGTCAGTATCTGGAACCTATAAAGCTAAACGATGTA 41533
337 337
41532 CGGTAAGCGTAAGTGTGTTTTTACCCCTTTTGTTCATCCTAGTTT 41483
338GlnValAspTrpLysSerMetAspLe 346
41482 CACTGGTTTCTCTATTTCATTCAATAAGGTTGACTGGAAGCAAGGACCT 41433
346 uSerTyrLeuLeuGlu..... 351
41432 GGGATACCTGACAGAGGTAAACATTTACAGAACTCGGCTTTTTTTTTTA 41383
351 351

41382 TACCAATTTTGCATGTTTGTAGTGGAGAAATGTAGTTTTTTAAACCTGTTCTC 41333
352AspAsnTyrVallyshisPheGlyAspLeuVally 363
41332 TCATGCTCATCAAGGAACTATACCAAGTACTTTTCTGCTAGTGAG 41283
363 slyAlaLysProIleHisGlyAlaAspAlaValLeuLysAlaPheAsnI 380
41282 ACAAGCAGCAACAATCAAGGTTCTGACCTTGCTTAAGGCTCAAAACA 41233
380 leAspGlyAspValargIleGlnTyrArgAspGlnLeuAspPheGluAsp 396
41232 TAAAGGATCATGTTCTGATCCGATATAAGACCAAGTAGAGTTTGAACGC 41183
397 IleAlaArgGlnPheGlyIlePheGluGlnTrpLys..... 408
41182 ATTGCAGGGAATTTGGTATATTTGAAGATGAAGGTAATTTGAGATG 41133
408 408
41132 ATTGAACATAATCTTCAATAATTTAGTCCAAATGGTCACATTTT 41083
409AspGlyValProArgAlaAlaTyrLysGly 418
41082 GTACCCCTTTGATGTTAGATGTTGTCCTCAACAGCATATAAGGA 41033
419 IleValValPheArgPheGlnThrSerArgArgValPheLeuValSerPr 435
41032 GTAGTGGTGTTCGAATCCACACAAGAGCGTGTATCTCTGGTGGGCC 40983
435 oAspSerLeuArgGlnLeuGlyValGluAspThr 446
40982 AGATTCGTAAATGACGCTTGAATTCGAAATTC 40949

seq_name: gb_pl:ATF22113

seq_documentation_block:

LOCUS ATF22113 93760 bp DNA linear PLN 27-AUG-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F22113 (ESSA project).

ACCESSION AL035539
VERSION AL035539.1 GI:4539331

SOURCE thale cress.
ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 93760)
AUTHORS Bevan,M., Wedler,H., Kutzner,M., Wambutt,R., Bancroft,I., Mewes,H.W., Mayer,K.F.X. and Schueller,C.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 93760)
AUTHORS EU Arabidopsis sequencing,project.

TITLE Direct Submission
JOURNAL Submitted (26-FEB-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de,Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosome 4 can be viewed at: <http://webserv.mips.biochem.mpg.de/proj/thal/>.

FEATURES

Location/Qualifiers

1..93760
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(2..106)
/gene="F22113.10"
/number=1

source

complement(1527..1619)
/gene="F22113.10"
/number=7
complement(1620..1857)
/gene="F22113.10"
/number=8
complement(1858..1900)

exon

complement(1436..1526)
/gene="F22113.10"
/number=6
complement(1331..1435)
/gene="F22113.10"
/number=5
complement(1269..1330)
/gene="F22113.10"
/number=4
complement(941..1088)
/gene="F22113.10"
/number=3
complement(638..867)
/gene="F22113.10"
/number=2
complement(362..568)
/gene="F22113.10"
/number=1
complement(107..190)
/gene="F22113.10"
/number=1
complement(191..361)
/gene="F22113.10"
/number=2
complement(362..568)
/gene="F22113.10"
/number=2
complement(569..637)
/gene="F22113.10"
/number=3
complement(638..867)
/gene="F22113.10"
/number=3
complement(868..940)
/gene="F22113.10"
/number=4
complement(941..1088)
/gene="F22113.10"
/number=4
complement(1089..1167)
/gene="F22113.10"
/number=5
complement(1168..1268)
/gene="F22113.10"
/number=5
complement(1269..1330)
/gene="F22113.10"
/number=6
complement(1331..1435)
/gene="F22113.10"
/number=6
complement(1436..1526)
/gene="F22113.10"
/number=7
complement(1527..1619)
/gene="F22113.10"
/number=7
complement(1620..1857)
/gene="F22113.10"
/number=8
complement(1858..1900)

intron

gene

CDS

gene

CDS

gene

CDS

```

/ gene="F22113.10"
/ number=8
complement(1901..1963)
/ gene="F22113.10"
/ number=9
complement(1964..2055)
/ gene="F22113.10"
/ number=9
complement(2056..2101)
/ gene="F22113.10"
/ number=10
complement(2102..2178)
/ gene="F22113.10"
/ number=10
complement(2179..2255)
/ gene="F22113.10"
/ number=11
complement(2256..2448)
/ gene="F22113.10"
/ number=11
complement(2449..2528)
/ gene="F22113.10"
/ number=12
complement(2529..2668)
/ gene="F22113.10"
/ number=12
complement(2669..2756)
/ gene="F22113.10"
/ number=13
complement(2757..2969)
/ gene="F22113.10"
/ number=13
complement(2970..3020)
/ gene="F22113.10"
/ number=14
complement(3021..3124)
/ gene="F22113.10"
/ number=14
complement(3125..3193)
/ gene="F22113.10"
/ number=15
3517..4827
/ gene="F22113.20"
/ number=1
3517..4827
/ gene="F22113.20"
/ number=1
3517..4827
/ gene="F22113.20"
/ number=1
3517..4827
/ note="similarity to amino acid transport protein -
Arabidopsis thaliana, PID:g2576363
contains EST gb:T43079, T20761, T41914, Aa586275, T43892"
/ codon_start=1
/ product="putative amino acid transport protein"
/ protein_id="CAB37481.1"
/ db_xref="GI:4539333"
/ translation="MGONEASSSSYTLKIPPPAREDSPLLGKPPPLSSQPKTFANVF
IAVAGVGLGYPYAFKRTGWLGMVLLIVSVSLTHHCMLLVYTRRKLDSPFNAGISKI
GSRDGLGAVGCSGRIVVDLFILSQAGFCVGLIFIGTTLANLSDPESPTSLRHQF
TRLGSEFGVSSKSLIWCFCFQGLNSKLTTHLAPLSIFADIVDLGAMAVVIED
SMILKQRPDVAFGMSLFYMGVAVYSPFEGVMVLPLESEMKKDKFKGVIALGM
GFILYTAFIGLGLAFGEDTMDIITANLGLAVSVTVOLGLCINLFTFPLMNPV
FEIVRRFSRGMYSAMLRLVLAIVLAVLPVNFADFLSVGSSTCCVLGLFVLPALE
HLVFKEMGHLWSSDTAIVLGLVLAIVSGTSSUSEIFSVKV"
complement(5061..5222)
/ gene="F22113.30"
/ number=1
complement(join(5061..5222,5303..5526,5567..5777,
5859..6034,6112..6127))
/ gene="F22113.30"
5061..6127
/ gene="F22113.30"
complement(join(5061..5222,5303..5526,5567..5777,
5859..6034,6112..6127))

```

```

/ gene="F22113.30"
/ note="similarity to gene T10 protein - mouse,
PIR2:S37488"
/ codon_start=1
/ product="putative protein"
/ protein_id="CAB37482.1"
/ db_xref="GI:4539334"
/ translation="MKITTEALRWEDGETVGGRLVGGTWNLCGTRHGRFLAFLTNFK
EASSFPAKSRGDLPLRYLQSEKSPAEAEETQDEISLYNGNLVVAHVLSKSMIYIT
NPPPHGDKLVTVQSPGIVHLSNANLSPKPNVSPNCVFLSLQCLRLREGFQOLLAEN
GSGEFPVKTMVEEVTNTVKDEETELPHVFTPTETVYHLSIFVDMORPTGRVGTSSIS
AIVKSHGDDGGDGEICFYERHLEEGDSNKEHTQQPVIIQNQSI"
complement(5223..5302)
/ gene="F22113.30"
/ number=1
complement(5303..5526)
/ gene="F22113.30"
/ number=2
alignment_scores:
Quality: 1131.50 Length: 1062
Ratio: 2.822 Gaps: 27
Percent Similarity: 37.759 Percent Identity: 31.262
alignment_block:
US-09-591-466C-2 x ATF22113/rev ..
Align seg 1/1 to reverse of: ATF22113 from: 1 to: 93760
9 AsLeuArgTyrLeuLeuValAlaAlaLeuAlaPheIleTyrIle.. 24
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3175 GACTTGAGATTCTTCTCATCCCGCAGCTTTCATGTTTCATACATCCA 3126
24 .....
3125 GGTCATCTATATCCGATGATCTCGCTTTTGAAGTTGAACATGAATG 3076
24 .....
3075 ATTCCATCGCTGCAACCTAAATCTGTCTGAATATATCTCTCTCTCTTC 3026
25 Gln..MetArgLeuPheAlaThrGlnSerGluTyrValAspArgLeuAla 40
::: |||||:|||||:|||||:|||||:|||||:|||||:|||||:
3025 TCTAGATAGGCTTTTCCAGACCAATCACAGTATGCAGATCGCCTCAGT 2976
41 AlaAla.....
:::||||
2975 TCCGCTGTGAGTTCTTACTTTTACCACCTGCCCTTTTGTCTTTAGACCATG 2926
42 .....
2925 CTACTAGTTCCTTTGAATCCCATTTTGTGGCTCTTTACAAGCCCATGA 2876
42 .....
2875 ACTACTTTCAAGAAGAAGACACTATCTGCACATTATACACCATATCATGCC 2826
42 .....
2825 CTTTGGTATTTCGATTTTTCCTGTGTGTGATGAAGATCTATACACAGC 2776
43 .....
2775 CTTTCATTTGTTTCTCCAGATCGAATCTGGAACCATTTGCACATAGTCAAA 2726
53 hrArgLeuLeuIleAspLysIleSerGlnGlnGlnIleValValAla 69
||| |||||:|||||:|||||:|||||:|||||:|||||:
2725 TCCGAGGCCTCATAGATGAAGTTAGCATCAAAACAGTCGCGGATGTTGCC 2676
70 LeuGlu..Glu.....
|||||
2675 CTCGAAGGTATATCCCTCTCTTCTTCTCCTCATTTGGGACTTGTGGAATCCT 2626

```

```
72 ..... 72
2625 AGATGACCATGTTCTTGGAACTATTGAATACTATACAAATAAATTT 2576
73 .....GlnM 74
2575 CTAAATGGAGTAGAAGACTTATACATTGATTCTGAGTGTCTTGAGTA 2526
74 eLysHis...GlnAspGlnGluCysArgGlnLeuArgAlaLeuValGln 89
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
2525 TGAAGAACCGCCAGGACGACGAAGAACTTTGGCAGCTTAAGGATCTAATCCAG 2476
90 AspLeuGluSer..... 93
: |||||
2475 ACGTTTGAAGTGGCTTACTATACCTGTAAAACTCATTTTCTTTGCTTTC 2426
94 .....Lys.Gly 95
||| |||
2425 TTTTATTAGTTTTTCTACTAATGCTTATTTTCTTCTCCAGAAAAAGGA 2376
96 IleLysLysLeuIle.....GlyAspVal..... 103
||| ||||| |||||
2375 ATAGCAAAACTCACTCAAGTGGACAGGCTCTGGATAAGAAGATGCCTTTC 2326
103 ..... 103
2325 TTCGTATTTGTAAGTGTATCACTTGTCTATATATATACATCTAACCTGTT 2276
104 .....GlnMetProValAlaAlaValValMetAla 114
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2275 TCTTGTTTTTATTAGCAGATCCCTGTGGCTGCTGTGAGTGGTTATGGCC 2226
115 CysSerArgThrAspTyrLeuGluArgThrIleLysSerIleLeu.... 129
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
2225 TGCAGTGTGCAGACTATCTTGAAGGACTGTGTAATCAGTTTAACTG 2176
129 ..... 129
2175 AGATACTACTGATTAAATCTTATACCACATTTTGATTGAACATATCTA 2126
130 .....Lys..TyrGlnThrSerValAlaSerLysT 139
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2125 ATGGCTATTCTTTTTTATACAGATATCAAACTCCGTTGCTTCAAAAT 2076
139 yrProLeuPheIle..... 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2075 ATCCTCTATTATATCTCAGTACTTAATTTGCAGGGTGGTGGTCTCTTA 2026
143 ..... 143
2025 GACTGTTTTATATTCACGAATTACTAACCATGGTCTTTTCCACGACAT 1976
144 ....Ser..GlnAspGlySerAsnProAspValArgLysLeuAlaLeuSe 158
||| ||||| |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1975 ATTTTCTGACAGGATGGATCTGTATCAAGCTGTCAAGAGCAAGTCAATCGAG 1926
158 rTyrGlyGlnLeuThrTyrMet.Gln..... 166
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1925 CTATAATCAATTAACATATATGCAAGGTAATTAATCTAGCGTATTTTC 1876
166 ..... 166
1875 TCCCTCAGTTCACAGTAGATAAATGAGGATGAAGCTCGTITTCCTCCCTTT 1826
167 ....HisLeuAspTyrGluProValHisThrGluArgProGlyGluLeuV 182
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1825 TCAGACTTGATTTTGAACAGTGGTCACTGAAAGGCCCTGGCGAACTGA 1776
182 alaIaTyrTyrLysIleAlaArg..... 189
: |||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
1775 CTGCGTACTACAAGATGCAGTAAGGACTGGTTTCTCTTTTCTCTGCGGT 1726
190 .....HisTyrLysTr 193
```



```

211 ..... 211
103352 ATAACATTTTCTTCTTGATCTGCTCTTTCCATCCCATTTATGGTA 103303
212 ..... AspAspMetGluIleAlaAlaAspPhePheAspTyr 223
103302 ATAATTTGCTAATAGATGATATGGAGATCGCCCTGATTTTTCAGCTAC 103253
224 PheGluAlaGlyAlaThrLeuLeuAspArgAspLys..... 235
103252 TTCGAGGCTGCAGCTAAATTTACTGATAACGATAAGTATGTTGATATTA 103203
235 ..... 235
103202 TATATTATTATAGTTGAAACCCCAAGGATTTCACTCTCCAATGATCTAA 103153
236 ..... SerIleMetAla 239
103152 TTGCTTTTGCTGCTATTCCTGTCACCTGTGGTATGAAGGACATCATGGC 103103
240 IleSerSerTrpAsnAspAsnGlyClnArgGlnPheValGlnAspPro.. 255
103102 GTTTCCTCCTGGAATGCAATGGACAAGCAATTTGTTACGACCCCAAG 103053
255 ..... 255
103052 TTAGTACTCTGATAGCCTGTATAAAGTTCTGTCTCCATTAATAGGA 103003
256 ..... AspAlaLeuTyrArgSe 261
103002 AAGTACTACTATTTTCCTACATAAACATTTTGCAGAAGCTCTTTACCGGTC 102953
261 rAspPheProGlyLeuGlyTrpMetLeuSerLysSerThrTrpSerG 278
102952 GGATTTCTTCTCCTGGACTGGATGGATGTTAACAAAGCCTACATGGATTG 102903
278 luLeuSerProLysTrpProLysAla..... 286
102902 AGCTGTCACTTAAGTGGCCCAAGCATATCTTTACTGTGCTACTAATT 102853
286 ..... 286
102852 TTTTATTAGCATATACAATATTTTCCATCTACCATATATTTCTTCCTTA 102803
287 ..... TyrTrpAspTrpLeuArgLeuLysGluAsnHis 299
102802 ACAAGACTTTACTTATTTGGATGACTGGTGAGGCTAAAGAGGTACACA 102753
299 rgGlyArgGlnPheIleArgProGlyValCysArgThrTyrAsnPheGly 315
102752 GAGACCGGCAATTTATTCGCCCGGAGTCTGCAGAACATACAATTTTGGC 102703
316 GluHis..... 317
102702 GAACATGTATGTTGGCTACTTCAATGCAAAATTAACCACATTTCTGC 102653
317 ..... 317
102652 ACACTAGTACCCTGCATAGTAATTTAGTCTTTGAGCCAGTAAGTATGG 102603
318 ..... GlySer 319
102602 ATGACTTGAATTTACACTAATAGATTTTCATACCTCCTTTTCAGGGGTCA 102553
320 SerLeuGlyClnPhePheLysGlnTyrLeuGluProIleLysLeuAsnAs 336
102552 AGCATGGGCAATTTTCAGACAGTACTTGGAAACCAATCAAGCTAAATGA 102503
336 pValGlnValAspTrpLysSerMet..... 344
102502 TGCCCATGTATGCAAACTTCCATCACCCTTTGTTTGTAACTGTGTC 102453
345 ..... AspLeuSerTyrLeuLeu..... 350

```

```

102452 AAATTCCTTTAGTCGACAGATTTACCATTTTACCTCTCTCCTTTGTGGC 102403
350 ..... 350
102402 TATAGATCAAGTGGAAATTCCTGAGGACCTGAGCTACCTCAAGGAGTAAGA 102353
350 ..... 350
102352 ATTTGTTTGGACTTTGAAACAGTAGTTTGCAGCAAAATCCCTAATATTAT 102303
351 ..... GluAspAsnTyrValLysHisPheGlyAspLeuValLys 363
102302 GTTGGCAACACACGACAGAAGTTCTGATCCAATTTGGAAAGACGTCGCT 102253
364 LysAlaLysProIleHisGlyAlaAlaAspAlaValLeuLysAlaPheAsnI 380
102252 AGTCCACCCCTCTCCATGGATCTGATGCCGCGTTGAAAGCCCAATAT 102203
380 eAspGlyAspValArgIleGlnTyrArgAspGlnLeuAspPheGluAspI 397
102202 GGATGCGGATGTAAGGATCCAGTATAACGATCAGGAAGACTTTTGAGCGGA 102153
397 leAlaArgGlnPheGlyIlePheGluTrpLys..... 408
102152 TAGCTCGTCAATTTGGAATATTTGAAGATGGAAGGTTTACTTCTTTTGAC 102103
408 ..... 408
102102 CCCATTTTATTCAATAAATGCTCTTGTCAAGTAGGTCAAGTGACAA 102053
408 ..... 408
102052 TTTTGTGCTGTGATTTTTCAGTGGTGATGTTAGTCTAGTAGAAAAATGT 102003
408 ..... 408
102002 TTCAGTTTGCAGTTTTCAGTCTAGATAGACATCAGCCCCGCGTAATGGTCTA 101953
408 ..... 408
101952 TAATTTTCTTTGGACTTCTCTGTTTTTTTAGTGTGTGATGATCATGTAT 101903
409 ..... AspGlyValProArgAlaAlaTyrLysG 418
101902 GTCTGTTTCTCTGTTTTCAGATGGCATTCAGAACAGCTTATAAAG 101853
418 lyIleValValPheArgPheGlnThrSer...ArgArgValPheLeuVal 433
101852 GAGTAGTAGTCTTCCGGTACAAGAGTAGTCAAGACGAATATACCTCGTT 101803
434 SerProAspSerLeuArgGlnLeuGlyVal 443
101802 GGTCGGATTCCTCAGTCAGCTTAGGGTA 101773

```